Service differentiation without prioritization in IEEE 802.11 WLANs

Suong H. Nguyen, *Student Member, IEEE,* Hai L. Vu, *Senior Member, IEEE,* and Lachlan L. H. Andrew, *Senior Member, IEEE*

***Abstract*—Wireless LANs carry a mixture of traffic, with different delay and throughput requirements. The usual way to provide low-delay services is to give priority to such traffic. However this creates an incentive for throughput sensitive traffic also to use this service, which degrades overall network performance. We show, analytically and by simulation, that the performance of both delay and throughput sensitive traffic can be improved by scaling IEEE 802.11's $CW_{\min}$ and *TXOP limit* parameters in equal proportion. This reduces, but does not eliminate, the incentive for bulk data users to use the low-delay service. We further show that this incentive can be removed, while still giving improved performance to both classes, by reducing the $CW_{\min}$ of the high throughput class by a constant that is independent of the traffic load.**

***Index Terms*—802.11 EDCA, service differentiation.**

I. Introduction

Wireless networks carry a diverse mix of traffic, from voice with tight delay constraints to bulk file downloads with only long-term throughput requirements. Efficient use of the network requires services tailored to each of these traffic classes. The traditional approach to providing quality of service (QoS) is to prioritize real-time traffic at the expense of data traffic, as done by the default parameter setting of the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) standard [1]. This creates an incentive for data applications to use the class intended for real-time users to gain a higher share of resources. This can degrade network performance drastically [17] and QoS differentiation no longer occurs when all data users use the highest priority class [15]. To cope with this, policing mechanisms have been proposed [16], which increase the complexity of the network.

As an alternative to prioritization, we propose a simple scheme that provides better service for both throughput- and delay-sensitive traffic, and encourages applications to use the service designed for them. The approach is based on the IEEE's standard 802.11e [1] for carrier sense multiple access with collision avoidance (CSMA/CA) using exponential backoff. Note that we only address "rational" users who try to optimize their performance without changing the network stack (e.g. application writers who optimize their code based on measured performance using all the available services). The issue of "rational" users who modify the wireless interface to gain a higher share of resources has been addressed in previous work. In particular, [2, 3] consider selfish users manipulating the backoff parameters and [4] considers those manipulating the clear channel assessment threshold. Besides, there have been several works addressing selfish users in a variety of other problems rather than service differentiation, such as rate control [5, 6], power control [7] and resource sharing [8].

The authors are with the Centre for Advanced Internet Architectures (CAIA), Faculty of Information and Communication Technologies, Swinburne University of Technology, Hawthorn, Melbourne VIC 3122, Australia. E-mail: {hsnguyen, hvu, landrew}@swin.edu.au.

TABLE I
Default MAC parameters in 802.11e EDCA (DSSS PHY)

AC	Traffic	$CW_{\min}$	CW_{max}	AIFSN	*TXOP limit*
AC_BE	Data	31	1023	3	1 packet
AC_VO	Real-time	7	15	2	3.264 ms

To support QoS, 802.11e uses EDCA, in which the access point (AP) selects four Access Categories (ACs) which stations can use. The ACs use different values of four MAC parameters: $CW_{\min}$, $CW_{\max}$, *TXOP limit* and $AIFS$. $CW_{\min}$ controls how long a station waits before transmission and *TXOP limit* controls how much it can transmit per channel access. Note that applications cannot choose arbitrary combinations of these parameters, but only those permitted by the AP.

The 802.11e EDCA standard recommends four particular combinations of parameters. The parameters for throughput-sensitive bulk data and delay-sensitive voice are shown in Table I, taken from Table 7-37 of [1]. In particular, the AC for real-time service is given higher priority than that for data service; every parameter is set to a more aggressive value. Because these recommended MAC parameters are static, they will not be good for all traffic loads. As a result, there have been several proposals to improve the performance of EDCA by adapting these parameters dynamically to network conditions. In particular, several works propose to tune CW based on collision rate [9, 10] or idle channel status [11]. Moreover, [12] tunes both CW_{min} and CW_{max} based on collision rate. Due to their adaptiveness, these schemes perform well when network load changes but they require collecting statistics such as collision rate or idle channel status frequently and updating CW accordingly. Note that these schemes are priority-based, which provide higher priority for realtime traffic. The priority-based QoS provision works well provided that the high-priority class is only used by low-throughput real-time traffic. However, when users are rational, this approach creates an incentive for all users to use the class AC_VO, resulting in no QoS differentiation and worse overall network performance.

In this paper we instead seek to provide service differentiation without prioritizing one class over another, that is, there is no ordering of the classes such that one gets better performance in all respects than the later ones. We do this by choosing ACs such that some parameters are less aggressive whenever others are more aggressive. Our aim is to provide "different but fair" service for different traffic types, by allowing users to choose different points on a throughput-delay tradeoff curve. Our proposal does not require any additional mechanisms such as fair queueing or traffic policing.

Note that fair differentiated service has already been proposed in [19–22] for wired networks. This has not been widely deployed, because it requires signals to be sent from the application to the core network. In contrast, for wireless links connected directly to the host running the application, no protocol changes are needed. Prior work in the wireless context can be divided into rewarding schemes [14, 15] and pricing schemes [17, 18]. The former approach uses 802.11e's contention-free period (CFP) to provide extra throughput to the data class, which is problematic because current wireless network interface cards do not implement the CFP. The pricing approach either requires micropayments of monetary prices, which makes implementation difficult, or must impose prices through some other form of service degradation such as packet drops, which seems counter-productive.

Our contributions are 1) propose the "proportional tradeoff" scheme which improves service for both traffic types compared to EDCA parameters (Section II), 2) provide a model of 802.11 WLANs with rational throughput-sensitive and delay-sensitive users (Section III) and show that there is still an incentive for data users to use the real-time class (Section IV), and 3) suggest a simple change to the proportional scheme (called the "proportional incentive adjusted" scheme or "PIA") to give throughput-sensitive applications the incentive to use the bulk-data service while giving improved performance to both classes (Section V).

II. Proposed proportional tradeoff scheme

We propose a mechanism which improves service for both data and real-time traffic by increasing $CW_{\min}$ and *TXOP limit*. We do not use the AIFS parameter because it provides load-dependent prioritization which makes it difficult to achieve a "fair" service differentiation.

In particular, we define $m > 1$ service classes, denoted by B_k $(k \in [1, m])$. These classes can cover different types of users with different requirements of delay and throughput. Users which demand higher throughput and can tolerate higher delay can transmit more packets per channel access but less often. To achieve this, class B_k with higher k has higher *TXOP limit* but commensurately higher $CW_{\min}$. This is similar to the method in [13] to ensure fairness.

Let $\mathcal{T}$ be the TXOP limit of class B_1, which is chosen to fit one packet at the lowest data rate supported by the standard. Then,

$$\text{TXOP limit of class } B_k = \eta_k \mathcal{T}, \tag{1a}$$

where η_k $(k = 1, \cdots, m)$ satisfies $\eta_k < \eta_{k+1}$ and $\eta_1 = 1$.

Let W_{B_k} be the value of $CW_{\min}$ used by class B_k. Then

$$W_{B_{k+1}} = \frac{\eta_{k+1}}{\eta_k} W_{B_k}. \tag{1b}$$

Our scheme provides several classes for different types of traffic; however, to remain simplicity, we only consider in this paper two extreme types of traffic: delay-sensitive and throughput-sensitive traffic. Note that class B_1 is designed for delay-sensitive traffic while class B_m is suitable for throughput-sensitive traffic. The logic is that real-time traffic requires low delay and often has only one packet to send at a time but the packet needs to be sent as soon as possible; hence, it always uses class B_1. In contrast, data sources require high throughput; hence, it may be willing to wait a little longer, if this increases the amount it can transmit per channel access.

We will show below when all data users use class B_k, their throughput improves when k increases. When η_k is appropriately chosen, this scheme improves service for both traffic types. This benefit comes from the reduction of collision probability in the network due to the lower attempt probability of data sources.

III. Model

Here we present a model of 802.11e EDCA WLANs with rational data and real-time users. Consider an infrastructure network with a set $\mathbb{S}$ of $N_s \geq 1$ saturated sources and a set $\mathbb{U}$ of $N_u \geq 0$ unsaturated Poisson sources with negligible queueing. (For a discussion of unsaturated sources with non-negligible queueing, see Appendix B-A2.) Non-saturated sources represent real-time users while data users are modeled as saturated sources. For simplicity, we make the standard assumption that each station transmits packets of only one source, although this is not required by the scheme itself. (For a discussion of multiple sources per station, see Appendix C.)

The natural framework for considering incentive issues is game theory. WLANs with rational users can be modeled as a game in which users are players. A player i chooses an action which is to use either of classes B_k-s. Based on other players' actions and its action, the player i will get a payoff, which is the throughput for a saturated user or the reciprocal of delay for an unsaturated user.

Using class B_1 is a dominant strategy (see [27]) for unsaturated stations, since it reduces their delay regardless of what other stations do. For this reason, we will not treat unsaturated stations as players, but simply model their effect on the throughput obtained by the saturated users.

In the model, s, s_k and u denote any saturated user, a saturated user using class B_k and an unsaturated user.

Let N_x $(x \in \{s, s_k, u\})$ denote the number of users of type x. Note that $N_s = \sum_{k=1}^{m} N_{s_k}$. Besides, let W_x $(x \in \{s_k, u\})$ be the minimum contention window of users of type x. Note that our model considers $W_x > 11$ to guarantee system stability as explained later in Section III-A1.

Different nodes may use different physical layer bit rates. To avoid inefficiencies [23], we aim at time fairness among saturated users, and so measure throughput as the amount of time each can transmit. In addition to the natural measure of the *fraction* of time (S_x, measured in seconds/second), some of our results apply to the more tractable measure of throughput in seconds/slot, denoted C_x. By *slot*, we mean MAC slot.

Our model makes the standard assumption that the network is in equilibrium. It also assumes that a saturated source sends data for the whole duration of TXOP limit. This is because a saturated source is defined as always having packets waiting to transmit.

A. Game Framework

A game of the wireless network described above is denoted by a quadruple $\langle \mathcal{P}, A, (u_i)_{i\in\mathcal{P}}, N_u \rangle$ where

- $\mathcal{P} = \{1, \ldots, N_s\}$, the set of players, contains the saturated users.
- For every $i \in \mathcal{P}$, $A = \{B_k : k \in [1, m]\}$ is the set of actions available to player i, where action B_k is to use MAC parameters $(CW_{\min}, TXOP) = (W_{B_k}, \eta_k \mathcal{T})$ ($W_{B_k} > W_{B_{k-1}}$ and $\eta_k > \eta_{k-1}$). Note that all the players have the same action space. However the game in Section IV has a different action space from that in Section V.
- For every $i \in \mathcal{P}$, the payoff $u_i(a)$ is the throughput of player i under the action profile a which is a vector containing the action of every player, $(a_1, \cdots, a_{N_s})$. There are two forms of the game, corresponding to the two types of throughput which are determined using the wireless model in Section III-A1 below.
  - Game 1: $u_i(a)$ is throughput in seconds/slot. Then, it is denoted by $C_i(a)$, given by (3);
  - Game 2: $u_i(a)$ is given by throughput in seconds/s. Then, it is denoted by $S_i(a)$, given by (4).

Note that this is a symmetric game [26], since each player has the same opportunities, and for each player, the same actions yield the same payoffs.

Our results use action profiles defined as follows

$$a_{(X;\cdot)} \in \{a \in A^{N_s} : a_1 = X\}, \quad \forall X \in A$$
$$a_{(X;\cdot;Z;\cdot)} \in \{a \in A^{N_s} : a_1 = X \text{ and } a_j = Z\}, \quad \forall X, Z \in A$$

1) Wireless model: We now present the wireless model to determine the throughput of a saturated station as payoff of a player in the game framework, which is derived, justified and validated in [24].

The model assumes that sources have no limit on the number of retransmission and $CW_{\max}$. This is made for notational and computational simplicity; however, simulations show that qualitative results from this model still hold when these two backoff parameters are truncated as in the standard.

Sources are indexed in non-increasing order of their packet duration. That is, $T_x \geq T_y$ for $x < y$.

The backoff mechanism imposes a slotted structure on time, with slot sizes denoted by a random variable Y. Note that Y is equal to σ if a slot is idle, to T_x if a slot contains a collision involving the source x and only sources $y > x$ with packets no larger than T_x, or to T_x^s if a slot contains a successful transmission of a source $x \in \mathbb{S} \cup \mathbb{U}$. Note that T_x is the time that a one-packet burst sent by a source x occupies the channel.

In each slot, source x ($x \in \mathbb{S} \cup \mathbb{U}$) attempts to transmit with "attempt probability" τ_x and, conditional on making an attempt, collides with "collision probability" p_x. Also let λ_u be the packet arrival rate of an unsaturated station $u \in \mathbb{U}$.

Central to the model is a set of fixed point equations. We only consider balanced fixed point, i.e., ones in which all the nodes of the same type have same value of collision probability, based on the following observations. The minimum contention window we consider is $W_x > 11$ ($x \in \{s_k, u\}$), for which binary backoff satisfies the condition of Theorem 5.4 in [25]; hence, the system has a unique fixed point which is balanced when $N_u = 0$. For $N_u > 0$, we assume that the load of unsaturated users is light enough that there again exists a unique and balanced fixed point as most analyses assume.

a) Fixed point model: The attempt probability τ_s of a saturated source s is the mean number of attempts per burst divided by the mean number of slots per burst, and satisfies

$$\frac{1}{\tau_s} = \frac{W_s}{2}\frac{1-p_s}{1-2p_s} + \frac{1}{2}. \tag{2a}$$

This is interpreted as one attempt per $1/\tau_s$ slots which is the average number of slots per backoff stage. Note that all saturated users using class B_k has the same CW_{min}, $W_s = W_{B_k}$ and hence, the same attempt probability and collision probability, denoted by τ_{s_k} and p_{s_k}, respectively.

Next, the attempt probability of an unsaturated source $u \in \mathbb{U}$ with the arrival rate λ_u is the mean number of attempts of u per second divided by the mean number of slots per second.

$$\tau_u = \frac{\lambda_u \sum_{j=0}^{\infty} p_u^j}{(1/\mathbb{E}[Y])} = \lambda \mathbb{E}[Y] \frac{1}{1-p_u}. \tag{2b}$$

Finally, the collision probability of source $x \in \mathbb{S} \cup \mathbb{U}$ is the probability at least another source also transmits when a source x transmits.

$$p_x = 1 - \frac{\prod_{i\in\mathbb{U}}(1-\tau_i)\prod_{j\in\mathbb{S}}(1-\tau_j)}{1-\tau_x}. \tag{2c}$$

b) Throughput of data users: The throughput in seconds/slot C_{s_k} of a saturated source of class B_k is given by the probability the source transmits successfully a burst in a slot multiplied by the duration it can transmit.

$$C_{s_k} = \tau_{s_k}(1-p_{s_k})\eta_k \mathcal{T}. \tag{3}$$

The throughput in seconds/s S_{s_k} of a saturated source of class B_k is given by the throughput in seconds/slot divided by the average duration of a slot.

$$S_{s_k} = \frac{C_{s_k}}{\mathbb{E}[Y]}. \tag{4}$$

Another measure called "relative throughput" is also used. This is the throughput of a saturated source under the given scheme divided by that under the scheme with no service differentiation ($\eta = 1$).

IV. Properties of Proportional Tradeoff Scheme

We now consider the first specific game in the foregoing framework, which is based on the proportional scheme to provide service differentiation. An alternative based on PIA will be considered in Section V.

Under the proportional scheme given by (1), the action space of the game is

$$A0 = \Big\{ \big(\eta_k W_{B_1}, \eta_k \mathcal{T}\big) : k \in [1, m] \Big\}, \quad \text{where } \eta_1 = 1$$

and $(\eta_k W_{B_1}, \eta_k \mathcal{T})$ are the MAC parameters of class B_k.

A. Theoretical results

The following results will be proved for unbounded retransmission and $CW_{\max}$ and some results are for networks with only data users. However, we will show by simulation they apply when these assumptions are relaxed.

1) Service differentiation property: We first show that the proportional scheme improves service for both data and real-time traffic by considering the network in which all users use the class designed for them in Theorems 1 and 3. In particular, all saturated sources use class $\mathrm{B}_{k>1}$ and all unsaturated sources use class B_1.

We start with Theorem 1, proven in Section VI-G, which states that, in a network without real-time users, when all data users uses class B_k with $\eta_k > 1$ under the proportional scheme, they will receive higher throughput than when there is no service differentiation ($\eta_k = 1$).

Theorem 1. *Consider the wireless model (2)–(4), in the game $\langle \mathcal{P}, A0, (S_i)_{i\in\mathcal{P}}, 0\rangle$ with all data users using the same class. The throughput in seconds/s of data users increase when they use class with higher η_k.*

The above theorem is based on the following lemma proven in Section VI-F.

Lemma 2. *Under the wireless model (2), in the game $\langle \mathcal{P}, A, (S_i)_{i\in\mathcal{P}}, 0\rangle$ with all data users using class B_k, the collision probability and attempt probability of all data users decrease with the increase of their $CW_{\min}$.*

This lemma suggests that under the proportional scheme, when data users use higher class (higher η_k), their $CW_{\min}$ increases. Therefore, their collision probability reduces, which explains for their throughput increase as stated in Theorem 1.

To show the benefit of the proportional scheme for both data and real-time users, we consider a simple network of mixed traffic in Theorem 3. This theorem is proven in Appendix A using the wireless model with (2a) simplified to

$$\tau_{s_k} = \frac{2}{W_{s_k}} \frac{1 - 2p_{s_k}}{1 - p_{s_k}}. \quad (5)$$

to keep the algebra tractable, assuming that $W_{s_k} \gg 1$.

Theorem 3. *Consider the wireless model (2)–(3) with (2a) replaced by (5), in game $\langle \mathcal{P}, A0, (S_i)_{i\in\mathcal{P}}, N_u\rangle$ with $N_u = N_s = N_{s_k} = 1$, $\max(T_u, T_s) < 2\mathcal{T}$, and $\lambda T_u \le 1$.*

- (T3-1) *The throughput in seconds/slot of the saturated station increases when $\eta_k \ge 1$ increases.*
- (T3-2) *The collision probability of the unsaturated station decreases when $\eta_k \ge 1$ increases.*

Note that (T3-1) is quite sensitive to modeling assumption and overhead duration.

Although the result in Theorem 1 is for scenarios with only data users and that in Theorem 3 is for a simple mixed-traffic scenario, we show by simulation that they hold for more general scenarios. In particular, simulation shows that the reduction in collision probability is accompanied by a reduction in the mean delay, except at light load.

2) Incentive property: Here we will investigate the incentive of bulk-data users under the proportional scheme by examining different actions of theirs in Theorems 4 and 6.

We start with the following theorem.

Theorem 4. *Under the wireless model (2)–(4), in the game $\langle \mathcal{P}, A0, (S_i)_{i\in\mathcal{P}}, N_u\rangle$ with $W_i > 11$, a data user using class B_1 has higher throughput than any other data user using any class $B_{k>1}$ in the same network. Specifically, $S_1(a_{(B_1;\cdot;B_{k\ge1};\cdot)}) \ge S_j(a_{(B_1;\cdot;B_{k\ge1};\cdot)})$.*

This theorem is proven in Section VI-B and based on the following lemma which is proven in Section VI-A.

Lemma 5. *Consider the wireless model (2), in the game $\langle \mathcal{P}, A, (S_i)_{i\in\mathcal{P}}, N_u\rangle$ with $N_{s_j} \ge 1$ and $N_{s_{j+i}} \ge 1$ ($i, j > 0$). If $W_{j+i} \ge W_j > 11$ then data users using class B_j have an attempt probability equal to or higher than those using class B_{j+i}, $\tau_{s_j} \ge \tau_{s_{j+i}}$. Moreover, if $W_{j+i} > W_j > 11$ then $\tau_{s_j} > \tau_{s_{j+i}}$.*

The following theorem proven in Section VI-D states that, regardless of the actions of other data users, the remaining user is better off by using class B_1.

Theorem 6. *Consider the wireless model based on (2)–(3) with (2a) replaced by (5), in the game $\langle \mathcal{P}, A0, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 11$. We have*

$$C_1(a_{(B_{k>1};\cdot)}) < C_1(a_{(B_1;\cdot)}). \quad (6)$$

Although the throughput in Theorem 6 is in seconds/slot, simulation demonstrates this result still holds for seconds/s. The proof of Theorem 6 is based on the following lemma.

Lemma 7. *Consider the wireless model based on (2)–(3) with (2a) replaced by (5), in the game $\langle \mathcal{P}, A, (u_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_{s_k} > W_u > 11$. Data user 1 has a higher attempt probability and other data users has lower attempt probability when data user 1 uses class B_1 than when it uses any class $B_{k>1}$.*

This lemma explains for the increase of the throughput of data user 1 when it uses class B_1 as stated in Theorem 6.

Note that an action profile is a *Nash equilibrium* if no player gets higher payoff by changing its action while others keep theirs unchanged [27]. From Theorem 6, the action profile with all data users using class B_1 is a unique Nash equilibrium. Then, according to Theorem 1, the throughput of a data user at Nash equilibrium is less than that when all data users use class $\mathrm{B}_{k>1}$. Section V will consider an improved scheme that avoids that issue.

From Theorem 6, using class B_1 is a dominant strategy, which means that regardless of actions of other users, a given user always get the highest throughput by using class B_1. Hence, even if the action space consists of mixed strategies [27] (i.e., randomly selecting a class from a given probability distribution), the action profile with all data users always using class B_1 is still a unique Nash equilibrium.

B. Simulation results and discussion

Recall that the properties of the proportional scheme in Sec. IV-A are proved for unbounded retransmission and

TABLE II
802.11G MAC AND PHY PARAMETERS

Parameter	Symbol	Value
Data bit rate	r_{data}	54 $Mbps$
Control bit rate	r_{ctrl}	1 $Mbps$
Basic rate		2 $Mbps$
PHYS header	T_{phys}	192 μs
MAC header	l_{mac}	288 $bits$
ACK packet	l_{ack}	112 $bits$
Slot time	σ	20 μs
SIFS	T_{sifs}	10 μs
DIFS	T_{difs}	50 μs
Retry limit		7

TABLE III
MAC PARAMETERS OF CLASSES B_1 AND B_2

Class	$CW_{\min}$	CW_{max}	AIFSN	*TXOP limit*
B_2	W_{B_2}	$2^5 W_{B_2}$	2	$\eta\mathcal{T}$
B_1	W_{B_1}	$2^5 W_{B_1}$	2	$\mathcal{T}$

$CW_{\max}$ and some of them are for a network with only data users. Herein we will use simulation (*ns-2.33* [28] [29]) to validate those in more general scenarios with both data and real-time users, and a limited number of retransmissions.

In simulated networks, unsaturated and saturated sources send packets to access point, using the user datagram protocol (UDP). Unsaturated sources have the same packet size and produce Poisson traffic of the same arrival rate. Saturated sources have the same packet size and receive CBR traffic faster than they can transmit. We use the 802.11g parameters in Table II. Note that the results also apply if not all users use the same data bit rate, or the network is based on the 802.11b.

For tractability, we only consider two classes ($k \in \{1, 2\}$). The MAC parameters specific to classes B_1 and B_2 in the proportional scheme are given in Table III with $W_{B_2} = \eta W_{B_1}$ and $\mathcal{T} = 0.72$ms.

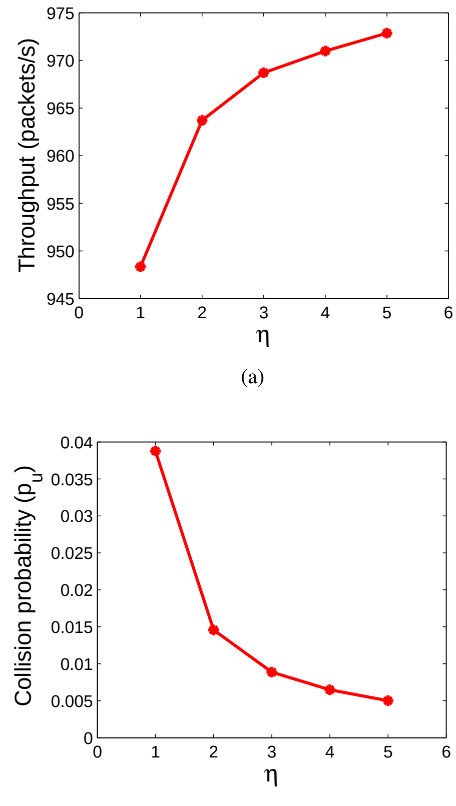
Note that the throughput in simulation results are measured in packets/s, which can be converted to seconds/s by multiplying by the packet duration. At 54 Mbps, this is 345 us for 1000 bytes, 375 us for 1200 bytes and 405 us for 1400 bytes.

1) Service differentiation property: To validate service differentiation property, we consider the network with all users using the class designed for them ($N_{s_1} = 0$). Realtime users use class B_1 and data users use class B_2.

a) Scenario 1 ($N_{s_2} = N_u = 1$): Fig. 1 shows the throughput of a data user and the collision probability of an unsaturated station. When η increases, the throughput increases and the collision probability decreases, which shows the benefit of the proportional scheme. This confirms the result of Theorem 3.

b) Scenario 2 ($N_{s_2} > 1$, $N_u > 1$): To investigate the ability of our scheme to give benefits to both classes of traffic in larger systems, we compare it with the default EDCA parameters (Table I) within the same scenarios.

The throughput of a data user and the mean delay of a real-time user under the proportional scheme are shown in Fig. 2, as functions of η for different N_s. Moreover, the performance metrics under the default setting (Table I) with all data users using class AC_BE and real-time users using class AC_VO and with all users using class AC_VO are also shown for comparison. In Fig. 2, the performance metric of



(a)

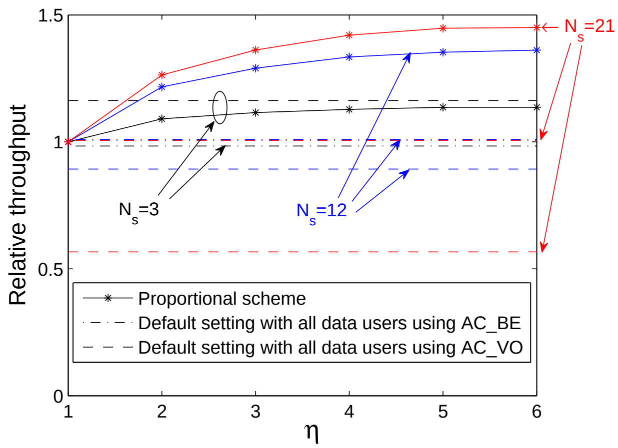
(b)

Fig. 1. Throughput of a data user and collision probability of a real-time user as a function of class B_2's *TXOP limit* in units of $\mathcal{T}$ (η). ($\lambda = 50$ packets/s, $l_{sat} = 1400$ bytes, $l_{nonsat} = 400$ bytes, $N_{s_2} = N_u = 1$, $N_{s_1} = 0$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1}$.)

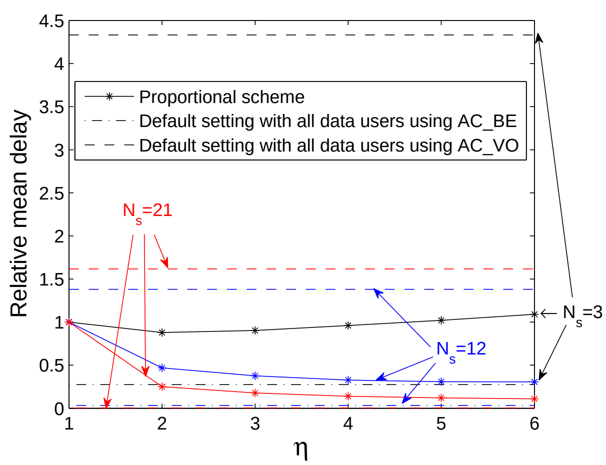
the proportional scheme at each η and the default setting are divided by that of the proportional scheme at $\eta = 1$. Note that the actual throughput and mean delay degrade as N_s increases; however, the relative performances improve as N_s increases.

Since the relative throughput is greater than 1 for $\eta > 1$ in Fig. 2(a), the proportional scheme with $\eta > 1$ always provides better service for data users compared to no service differentiation ($\eta = 1$). This corroborates the result of Theorem 1. Note that the benefit of the proposed scheme increases with contention level in the network. Fig. 2(a) also shows that the throughput of a data user in the proportional scheme is always higher than that in the default EDCA scheme with all data users using class AC_BE. Moreover, when traffic load is high enough, our scheme significantly improves the throughput of data users compared to the default EDCA setting with all data users using class AC_VO. At light traffic load (e.g. $N_s = 3$), our scheme provides slightly lower throughput for data users but significantly lower mean delay for realtime users than EDCA scheme with all data users using class AC_VO.

In Fig. 2(b), when the load is high enough, our scheme with $\eta > 1$ provides significant improvement in mean delay of real-time users compared to the case of no service differentiation ($\eta = 1$). At light load (e.g. $N_s = 3$), the improvement is negligible. This is acceptable because delay only becomes a problem at high load. Fig. 2(b) also suggests that at each network load, there exists an optimal value of η at which mean delay is minimum (e.g. $\eta = 2$ at $N_s = 3$ and $\eta = 5$ at



(a) Data throughput. Proportional allocation gives higher throughput than the default setting with data users using AC_BE class. For heavy load, it also gives better throughput than the default with data users using AC_VO.



(b) Real-time delay. Proportional allocation gives lower delay than the default setting with data users using AC_VO, though higher delay than the default with data users using AC_BE.

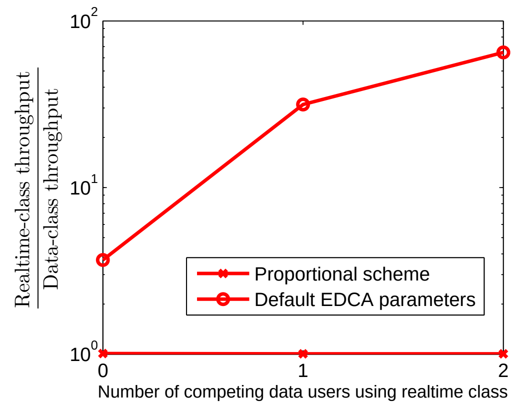
Fig. 2. Performance of proportional allocation as a function of class B_2's *TXOP limit* in units of $\mathcal{T}$ (η). (λ = 35 packets/s, l_{sat} = 1000 bytes, $l_{nonsat} = 200$ bytes, $N_u = 6$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1}$.)

$N_s = 12$). This optimal η increases with the network load.

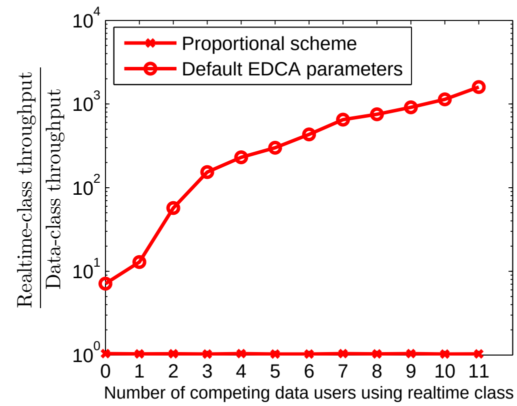
Compared with the default parameter setting which prioritizes real-time traffic with all data users using class AC_BE, we expect the performance will be worse for real-time users under the proportional scheme. This is seen in Fig. 2(b). However, compared with the default setting with all data users using class AC_VO, our proportional scheme provides much better service for real-time users.

Although the optimal η in our proportional scheme depends on traffic load, the majority of the benefit for both data and realtime users is obtained at $\eta = 2$. Fig. 2 suggests that increasing η beyond 6 or 7 does not improve performance significantly, which is because the contention level does not decrease much further then.

2) Incentive property: Here we will investigate the incentive of data users in choosing a class under the proportional tradeoff scheme, by comparing the payoff of a particular data



(a) $N_s = 3$



(b) $N_s = 12$

Fig. 3. Ratio of throughput of a data user when it uses "real-time" class and "bulk data" class as a function of the number of competing data users using realtime class. The figures show there is a big incentive for data users to use the realtime class under the default EDCA parameters while this incentive seems negligible under the proportional scheme. (λ = 35 pkts/s, l_{sat} = 1000B, $l_{nonsat} = 200$B, $N_u = 6$, $\eta = 2$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1}$.)

user in different action profiles. We assume realtime users always choose class B_1.

a) A class-B_1 user has higher throughput than a class-B_2 user: We simulated the network scenario with $\lambda = 35$ packets/s, $l_{sat} = 1000$ bytes, $l_{nonsat} = 200$ bytes, $N_u = 6$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1}$, η varied from 2 to 5, and $N_s = \{3, 12, 21\}$. What we have found is that a data user using class B_1 gains higher throughput than another data user using class B_2, which confirms the result of Theorem 4.

b) Nash equilibrium: Our results in Fig. 3 show that a data user achieves higher throughput by using class B_1 regardless of the other data users' choice under the proportional scheme. However, a data user has less incentive to use B_1 in this case than it does to use AC_VO under the default EDCA scheme, because the latter provides a larger increase in throughput relative to AC_BE.

This implies that the action profile in which all data users use the realtime class B_1 is the only Nash equilibrium, which confirms the result of Theorem 6. However, this equilibrium gives a lower throughput than could be obtained when all data users use class B_2, as shown by the increase in relative throughput with η in Fig. 2(a). We next investigate a way to avoid this undesirable equilibrium.

V. Incentive adjusted scheme

Section IV-B showed that for networks with both data and realtime users, our proportional scheme can improve service for both traffic types relative to the scheme with no service differentiation, especially at high load. However, when a small fraction of data users uses class B_1, their throughput can be slightly improved. Although the improvement is small, measurement-driven application design will still result in class B_1 being chosen by throughput-sensitive applications. However, we will now show a slight modification to the proposed scheme which can eliminate this incentive issue. This is in contrast to priority-based schemes, which require explicit policing or pricing mechanisms [14, 15, 17, 18].

A. Description of the incentive adjusted scheme, PIA

We modify the proportional scheme by reducing $CW_{\min}$ of class $B_{k>1}$ by an amount $\epsilon_k > 0$, which provides higher benefit for users of class $B_{k>j}$ than users of class B_j. The reduction in $CW_{\min}$ for class $B_{k>1}$ results in more throughput for a data user when it uses $B_{k>1}$ compared to when it uses class B_1, and thus data users have no incentive to use the real-time class B_1 but have incentive to use the class providing the highest throughput B_m. Recall that users can only select one of the access classes determined by the access point, and cannot choose arbitrary combinations of parameters.

Note that the performance of delay-sensitive users degrades as ϵ_k increases, and so we would like to use the smallest ϵ_k such that bulk data users using class B_k get a higher throughput than those using class B_{k-1}, regardless of the network load; any larger value of ϵ_k will increase that benefit but degrade realtime sources' performance. The absolute smallest such ϵ_k is given in Theorem 8. Importantly, it depends only on η_k and η_{k-1}, and not the number of users of each type in the network.

Theorem 8. *Under the wireless model based on (2)–(4), in the game $\langle \mathcal{P}, A, (S_i)_{i\in\mathcal{P}}, N_u\rangle$ with $W_{s_k} = \frac{\eta_k}{\eta_{k-1}} W_{s_{k-1}} - \epsilon_k > W_{s_{k-1}} > 11$, when*

$$\epsilon_k \geq \epsilon_k^0 = 4(\frac{\eta_k}{\eta_{k-1}} - 1), \tag{7}$$

data users using B_k get higher throughput than those using B_{k-1}. That is, $S_1(a_{(B_k;\cdot;B_{k-1};\cdot)}) > S_j(a_{(B_k;\cdot;B_{k-1};\cdot)})$.

The above result is proved in Section VI-E.

Specifically, under the incentive adjusted scheme PIA, the action space in the game framework has the form

$$A1 = \left\{ (W_{B_1}, \mathcal{T}), \left\{ \left(\frac{\eta_k}{\eta_{k-1}} W_{B_{k-1}} - \epsilon_k^0, \eta_k \mathcal{T} \right) \right\}_{k\in[2,m]} \right\}, \tag{8}$$

where $(W_{B_1}, \mathcal{T})$ and $\left(\frac{\eta_k}{\eta_{k-1}} W_{B_{k-1}} - \epsilon_k^0, \eta_k \mathcal{T}\right)$, respectively, are MAC parameters of class B_1 and $B_{k>1}$.

B. Properties of PIA

In this section, we first use the game framework above to derive some properties of the PIA scheme. Then, we validate these results using ns-2 simulation.

1) Theoretical results: Here the results will be proved for unbounded retransmission $CW_{\max}$ and some results are for networks with only data users. However, simulation shows they still apply when these assumptions are relaxed.

a) Service differentiation property: To show that PIA improves service for data traffic, we consider the network in which all users use the class designed for them in Theorem 9. It states that PIA provides better service for data users at classes with higher η_k, which is proven in Section VI-G.

Theorem 9. *Consider the wireless model (2)–(4), in the game $\langle \mathcal{P}, A1, (S_i)_{i\in\mathcal{P}}, 0\rangle$ with all data users using the same class. The throughput in seconds/s of data users increase when they use class with higher η_k.*

The following corollary comes from the above theorem.

Corollary 10. *Consider the wireless model based on (2)–(4), in the game $\langle \mathcal{P}, A1, (S_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 11$ and all data users using class B_k. The throughput in seconds/s of each data user using class $B_{k>1}$ under PIA is higher than that under no service differentiation (all use class B_1).*

b) Incentive property: To see if PIA eliminates the incentive for data users to use the realtime class B_1, we look at their performance under different actions.

The following theorem, proven in Section VI-I, implies that the action profile with all data users using the highest class B_m is the unique Nash equilibrium.

Theorem 11. *Consider the wireless model based on (2)–(3) with (2a) replaced by (5), in the game $\langle \mathcal{P}, A1, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 11$. For any action profiles in which not all data users use class B_m, a data user using a class other than B_m can improve its throughput by using B_m.*

Note that this theorem is a natural consequence of Lemma 12. This lemma, proved in Section VI-H, states that, if there exists at least another data user using the class with the index equal to or higher than the class used by a given user, the given user can get a higher throughput per slot by using the highest class, B_m.

Lemma 12. *Consider the wireless model based on (2)–(3) with (2a) replaced by (5), in the game $\langle \mathcal{P}, A1, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 11$. For all $i \geq 0$,*

$$C_1(a_{(B_k;\cdot;B_{k+i};\cdot)}) < C_1(a_{(B_m;\cdot;B_{k+i};\cdot)}). \tag{9}$$

Although the throughput in Lemma 12 is in seconds/slot, simulation shows that this result still holds for seconds/second.

Under PIA, the action profile with all data users using the highest class B_m is the unique Nash equilibrium. Then, according to Corollary 10, the throughput of each data user at this Nash equilibrium is greater than that when all data users use class B_1. This suggests that PIA achieves the desired goal of providing a scheme in which rational users will all gain improved performance.

Note that, when mixed strategies are allowed, it remains an open question whether the equilibrium in which all users use class B_m is the unique Nash equilibrium.

TABLE IV
MAC PARAMETERS OF CLASSES B_1, B_2 AND B_3

Class (B_k)	*TXOP*	ϵ_k^0	$CW_{\min}$	$CW_{\max}$	AIFSN
B_1	$\mathcal{T}$	0	W_{B_1}	$2^5 W_{B_1}$	2
B_2	$2\mathcal{T}$	4	$W_{B_2} = 2W_{B_1} - 4$	$2^5 W_{B_2}$	2
B_3	$3\mathcal{T}$	2	$W_{B_3} = \frac{3}{2}W_{B_2} - 2$	$2^5 W_{B_3}$	2

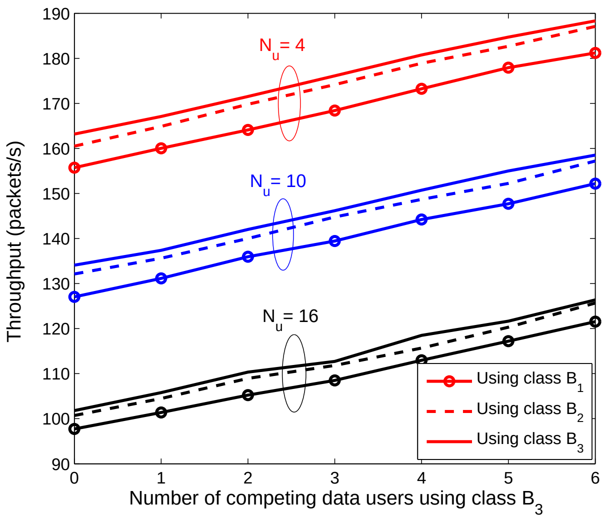


Fig. 4. Throughput in packets/s of a data user when it uses classes B_1, B_2 and B_3 as a function of the number of competing data users using data class B_3. The throughput improvement of a data source under the PIA scheme at a given N_u is the ratio of the throughput when all data users use class B_3 and the throughput when all data users use class B_1, which is about 22% at $N_u = 4$ and larger when N_u increases. ($\lambda = 30$ pkts/s, $l_{sat} = 1400$B, $l_{nonsat} = 300$B, $N_s = 8$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1} - \epsilon^0$, $\eta = 2$.)

2) Simulation results and discussion: Recall that the properties of PIA are proved for networks with only data users and for unbounded retransmission and $CW_{\max}$. Here we will use simulation (*ns-2.33* [28] [29]) to validate those in more general scenarios with both data and real-time users and a limited number of retransmissions. Noticeably, simulation results show that PIA is actually incentive-compatible, which means that using class B_m is the best strategy regardless of what other data users choose.

The simulated network in this section is the same as one in Sec. IV-B. Note that all the results are still valid when 802.11b parameters are used.

a) Incentive property: We verify here that throughput-sensitive users have an incentive to choose class with the highest TXOP. We consider the case of three ACs per station: B_1, B_2 and B_3. The MAC parameters specific to these ACs are given in Table IV with $\mathcal{T} = 0.72$ms.

Figure 4 displays the throughput of a data user at different choices of a class when other data users arbitrarily choose class B_1 or B_3. The results show that a data user obtains higher throughput by using class B_3 than by using B_1 or B_2, regardless of the choice between B_1 and B_3 of other users. This validates the result of Theorem 11. The total throughput is maximum when all data users choose class B_3 which is about 22% higher than the case when they all choose class B_1 at $N_u = 4$. This ratio becomes larger when N_u increases.

Note that the results for the case of two ACs per station (e.g. only B_1 and B_3 classes are available) can also be inferred from Figure 4. In particular, a data user always gets higher throughput by using class B_3, regardless of the choice of other users. This suggests that PIA is incentive compatible, resulting in all data users to choose class B_3. This property of PIA is even stronger than the one proven in Theorem 11. Note that the throughput improvements reported here are at the MAC layer only and without considering any congestion control mechanism of the higher layers.

b) Comparison with the default EDCA parameters: We can now compare the performance of the PIA with that of the default QoS classes (Table I), under the assumption that all users will use whatever class gives them the best performance.

We consider the case of two ACs per station: B_1 and B_2. The MAC parameters specific to classes B_1 and B_2 are given in Table III with $W_{B_2} = \eta W_{B_1} - \epsilon^0$, $\epsilon^0 = 4(\eta - 1)$, and $\mathcal{T} = 0.72$ms. In the scenarios considered here, the value of η is varied from 1 to 5. The case of no service differentiation ($\eta = 1$) is included for comparison.

Note that under the default parameters, all users will use AC_VO, and under PIA, bulk data users will use class B_2 and real-time users will use class B_1.

The relative throughput of a saturated user under PIA is shown in Fig. 5 as functions of η for different numbers of saturated users, N_s. For comparison, the throughput under the default parameter setting (Table I) is also shown. The throughput is again normalized by that of PIA at $\eta = 1$.

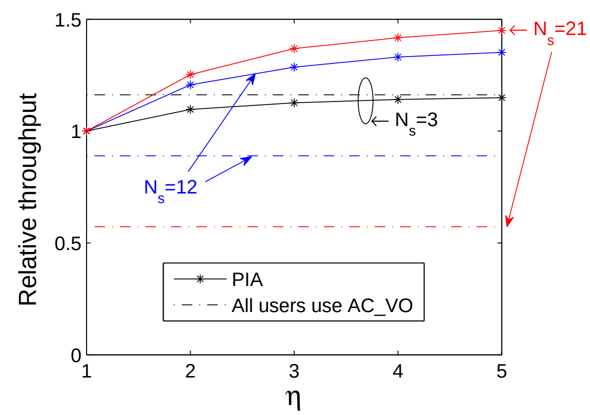
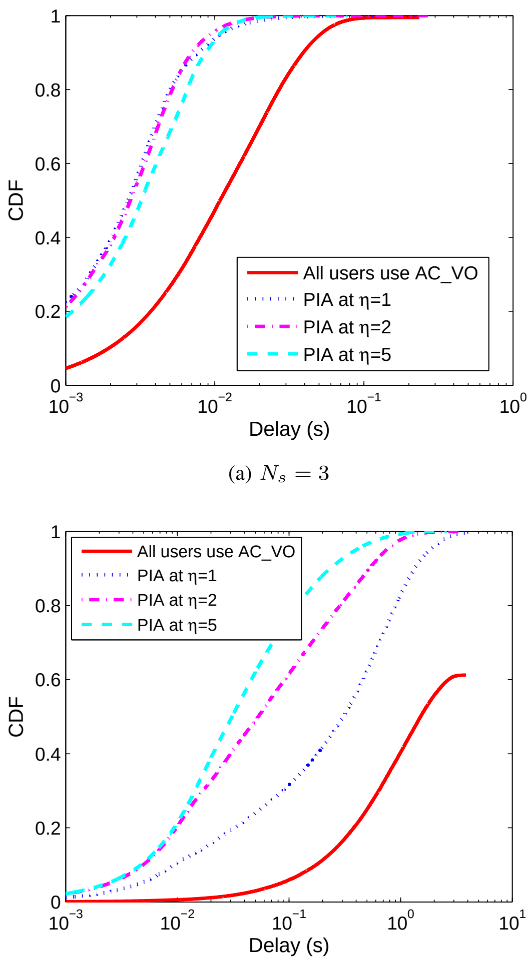


Fig. 5. Throughput of a data user under PIA as a function of class B_2's *TXOP limit* in units of $\mathcal{T}$ (η), scaled by that of PIA at $\eta = 1$. PIA gives better throughput than the default setting with data users using AC_VO class, except at light load. ($\lambda = 35$ packets/s, $l_{sat} = 1000$ bytes, $l_{nonsat} = 200$ bytes, $N_u = 6$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1} - \epsilon^0$ for $\eta = \{1, 2, 3, 4, 5\}$.)

Figs. 5 and 2(a) show that the throughput increases faster with η under PIA than it did under the proportional scheme, which reflects the reduction in $CW_{\min}$. This implies that PIA provides better service for data users than the proportional scheme without service differentiation ($\eta = 1$), especially at high load. This is in contrast to the default parameter setting with all data users using real-time class (AC_VO), for which the performance degrades rapidly at high load. For low load ($N_s = 3$), the default setting performs better than our PIA scheme because the more aggressive choice of $CW_{\min}$ is better matched to a small number of stations. However, the tradeoff is that the delay performance of realtime users under the default setting is worse than that under PIA.

This improvement in throughput of PIA comes at the expense of increased delay for real-time users. Figs. 6(a) and 6(b) show the probability that a packet of a real-time user is successfully transmitted before a given delay, for different η and loads $N_s = 3$ and $N_s = 21$.

Fig. 6(a) shows that PIA at both $\eta = 2$ and $\eta = 5$ gives a higher probability that a packet is successfully delivered at a given delay than the default setting with all data users using the class AC_VO. This means that the average packet delay under PIA is smaller. In this lightly loaded case, $\eta = 2$ provides comparable service to $\eta = 1$ (no service differentiation), and $\eta = 5$ provides slight degradation compared to $\eta = 1$, but less than that caused by the default prioritization setting.



(a) $N_s = 3$

(b) $N_s = 21$

Fig. 6. Probability a packet of real-time users is successfully delivered as a function of delay. ($\lambda = 35$ pkts/s, $l_{sat} = 1000$B, $l_{nonsat} = 200$B, $N_u = 6$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1} - \epsilon^0$ for $\eta = \{1, 2, 5\}$.)

In the heavily loaded case of Fig. 6(b), the cumulative distribution of delay for the default setting never reaches 1, which indicates a high loss rate. In contrast, PIA has a low loss rate for all values of η tested, although some packets have very high delays. In this case, the benefit increases as η increases. Together with the result in Fig. 6(a), this implies that under PIA, the optimal η for real-time users increases with traffic load, as was observed for the proportional scheme. However, even using $\eta = 2$ for all loads appears to provide improvement over the default parameters.

In brief, although the optimal η in PIA depends on traffic load, it is clear that when $\eta = 2$, PIA provides better service for both traffic types under typical scenarios considered in this section and Appendix B. This implies that when designing a network with an unknown number of users, PIA can be implemented by simply setting $\eta = 2$ and $\epsilon = \epsilon^0 = 4$. Adaptive schemes that set η dependent on the estimated load are possible, but out of the scope of this paper.

VI. Proofs

In this section, the proofs of all lemmas and theorems in Sections IV-A and V except Theorem 3 are shown.

We first express the mean slot time $\mathbb{E}[Y]$ in (4) as follows. Define G by

$$G = \prod_{i\in\mathbb{U}}(1-\tau_i)\prod_{j\in\mathbb{S}}(1-\tau_j) \tag{10}$$

The mean slot time $\mathbb{E}[Y]$ can be expressed in terms of the probabilities P^i, P^s_x and P^c_x that a given slot contains (a) no transmissions, (b) a successful burst transmission from source x, or (c) a collision involving the source x and only sources $y > x$ with packets no larger than T_x. Specifically,

$$\mathbb{E}[Y] = P^i\sigma + \sum_{x\in\mathbb{S}\cup\mathbb{U}} P^s_x T^s_x + \sum_{x\in\mathbb{S}\cup\mathbb{U}} T_x P^c_x \tag{11a}$$

$$P^i = G \tag{11b}$$

$$P^s_x = \frac{\tau_x}{1-\tau_x}G \tag{11c}$$

$$P^c_x = \frac{\tau_x}{1-\tau_x}\left(\prod_{y\le x}(1-\tau_y) - G\right) \tag{11d}$$

Durations T^s_x and T_x of source x using class B_k are

$$T^s_x = E + \eta_k\mathcal{T}, \qquad \text{with} \quad E > \sigma \tag{12}$$

and

$$T_x = E + T_{\mathrm{p}x} + T_{\mathrm{sifs}} + T_{\mathrm{ack}} \tag{13}$$

where E is the interval during which a station needs to sense channel free before transmitting (e.g. AIFS or DIFS). T_{sifs}, and T_{ack} are the durations of SIFS and an ACK packet, and $T_{\mathrm{p}x}$ is the transmission time of a packet from the source x.

A. Lemma 5

Proof: From (2a), we have

$$\frac{1}{1-p_{s_k}} = 2 - \frac{W_{B_k}}{2/\tau_{s_k} - 1}. \tag{14}$$

Next, dividing $1 - p_{s_{j+i}}$ from (2c) by $1 - p_{s_j}$ from (2c) gives

$$\frac{1-p_{s_{j+i}}}{1-p_{s_j}} = \frac{1-\tau_{s_j}}{1-\tau_{s_{j+i}}}. \tag{15}$$

To simplify notation, define

$$g(\tau, W) = \frac{1-\tau}{2-\frac{W}{2/\tau-1}} = \frac{1}{2}\frac{1-\tau}{1-\frac{W}{2}\frac{1}{2-\tau}\tau}. \tag{16}$$

Substituting $1-p_{s_{j+i}}$ and $1-p_{s_j}$ from (14) into (15) gives

$$g(\tau_{s_{j+i}}, W_{j+i}) = g(\tau_{s_j}, W_j). \quad (17)$$

Since $W_{s_k} > 4$ by hypothesis and $\tau \leq 1$, the coefficient $\frac{W}{2}\frac{1}{2-\tau}$ of τ in the denominator of (16) is greater than 1 and increasing in τ. Hence, $g(\tau, W)$ is increasing in τ. Moreover, $g(\tau, W)$ is increasing in W. Therefore, from (17), $W_{j+i} > W_j$ implies $\tau_{s_{j+i}} < \tau_{s_j}$ and $W_{j+i} \geq W_j$ implies $\tau_{s_{j+i}} \leq \tau_{s_j}$. ■

B. Theorem 4

Proof: Under the action profile $a_{(B_1;\cdot;B_{k\geq1};\cdot)}$, we have $N_{s_k} \geq 1$, $N_{s_1} \geq 1$, $S_1(a_{(B_1;\cdot;B_{k\geq1};\cdot)}) = S_{s_1}$ and $S_j(a_{(B_1;\cdot;B_{k\geq1};\cdot)}) = S_{s_{k\geq1}}$. Thus it is required to show $S_{s_1} \geq S_{s_k}$ under (2)–(4), with strict inequality if $\eta_k > 1$.

Dividing S_{s_k} by S_{s_1} from (4) and substituting (15) gives

$$\frac{S_{s_k}}{S_{s_1}} = \frac{\tau_{s_k}(1-p_{s_k})\eta_k}{\tau_{s_1}(1-p_{s_1})} = \frac{\tau_{s_k}(1-\tau_{s_1})\eta_k}{\tau_{s_1}(1-\tau_{s_k})}. \quad (18)$$

To show $S_{s_1} \geq S_{s_k}$ it is sufficient to show that the denominator of (18) is at least as large as its numerator.

First, by (2a) and the fact that $W_{B_k} = \eta_k W_{B_1}$,

$$\tau_{s_1}(1-\tau_{s_k}) - \tau_{s_k}(1-\tau_{s_1})\eta_k = \tau_{s_k}\tau_{s_1}(\frac{1}{\tau_{s_k}} - \frac{\eta_k}{\tau_{s_1}} + \eta_k - 1)$$
$$= \tau_{s_k}\tau_{s_1}\Big(\frac{\eta_k W_{B_1}}{2}(\frac{1-p_{s_k}}{1-2p_{s_k}} - \frac{1-p_{s_1}}{1-2p_{s_1}}) + \frac{\eta_k - 1}{2}\Big) \quad (19)$$

To show that (19) is non-negative, it is sufficient to show that $\frac{1-p_{s_k}}{1-2p_{s_k}} \geq \frac{1-p_{s_1}}{1-2p_{s_1}}$, or equivalently that $p_{s_k} \geq p_{s_1}$, since $p_{s_1} \geq 0$.

Under the action space $A0$ and by hypothesis, $W_{B_k} = \eta_k W_{B_1} \geq W_{B_1} > 4$, which satisfies the conditions of Lemma 5. Hence $\tau_{s_1} \geq \tau_{s_k}$, and by (15), $p_{s_k} \geq p_{s_1}$. If $\eta_k > 1$, these inequalities are all strict. ■

C. Lemma 7

Proof: To see how the attempt probability of the user 1 changes when its action changes from B_k to B_{k+i} ($i > 0$), consider an arbitrary action profile of the form $a_{(X;)}$ for some $X \in A$. Then there are a $j \neq 1$ and a c which depends on X and a_j, such that

$$W_1 = cW_j \quad (20)$$

By hypothesis, $W_1 > 11$, whence $cW_j > 11$. Note that subscripts 1, i and i in this proof are to denote the quantities for user 1, i and j.

We first prove that there exists a unique solution of the fixed point model and find that solution. We then show how the solution changes with the action choice of user 1.

Since $N_u = 0$ by hypothesis, (2c) implies

$$p_i = 1 - \frac{\prod_{k=1}^{N_s}(1-\tau_k)}{1-\tau_i}, \quad \forall i \in \mathcal{P}. \quad (21)$$

whence

$$(1-p_i)(1-\tau_i) = (1-p_j)(1-\tau_j), \quad \forall i \neq j \quad (22)$$

From (5),

$$p_i = 1 - \frac{2}{4 - W_i\tau_i}, \quad \forall i \in \mathcal{P}. \quad (23)$$

Replacing $1-p_j$ and $1-p_i$ from (23) into (22) gives

$$\frac{1-\tau_i}{4-W_i\tau_i} = \frac{1-\tau_j}{4-W_j\tau_j}. \quad (24)$$

This is equivalent to

$$\tau_i = \frac{(4-W_j)\tau_j}{4 - W_i + (W_i - W_j)\tau_j}, \quad \forall i \in \mathcal{P} \setminus \{j\}. \quad (25)$$

Substituting (20) and p_1 from (21) into τ_1 from (5) gives

$$\tau_1 = \frac{2}{cW_j}\left(2 - \frac{1}{\prod_{i=2}^{N_s}(1-\tau_i)}\right) \equiv f_1(\tau_j, c). \quad (26)$$

Note that f_1 is a function of τ_j due to the relation between τ_i-s in the denominator and τ_j given in (25).

Substituting $1-p_j$ from (23) into $1-p_j$ from (21) gives

$$\tau_1 = 1 - \frac{2(1-\tau_j)}{(4-W_j\tau_j)\prod_{i=2}^{N_s}(1-\tau_i)} \equiv f_2(\tau_j). \quad (27)$$

Then, a solution of the fixed point model is any solution to $f_1(\tau_j, c) = f_2(\tau_j)$ with $\tau_j \in [0,1]$. We first prove there exists such a solution and then prove its uniqueness.

Now $f_1(\tau_j, c)$ and $f_2(\tau_j)$ are decreasing functions of τ_j on $[0, 2/W_j]$, as illustrated in Fig. 7.

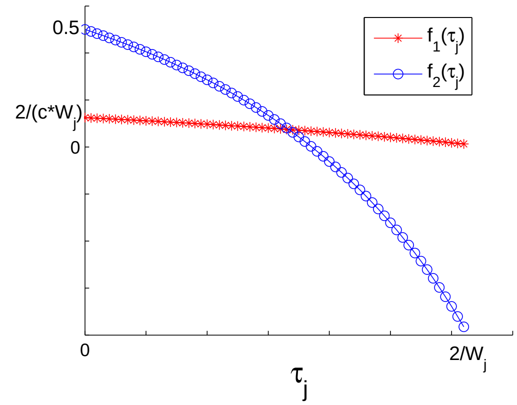


Fig. 7. $f_1(\tau_j, c)$ and $f_2(\tau_j)$

Moreover, at $\tau_j = 0$, we have $f_2(0) > f_1(0, c) > 0$. Besides, let τ_j^* be the solution to $f_1(\tau_j) = 0$. Then, (26) implies $\prod_{i=2}^{N_s}(1-\tau_i) = 1/2$. Substituting this into (27) gives

$$f_2(\tau_j) = 1 - \frac{2(1-\tau_j)}{(4-W_j\tau_j)(1/2)} = \frac{(4-W_j)\tau_j}{4-W_j\tau_j} < 0 \quad (28)$$

If τ_j^* is in (0,1) and unique, these, together with the continuity of $f_1(\tau_j, c)$ and $f_2(\tau_j)$, imply that there exists a solution to $f_1(\tau_j, c) = f_2(\tau_j)$ with $\tau_j \in (0, \tau_j^*)$. The following proves that τ_j^* is unique solution in (0,1) of $\prod_{i=2}^{N_s}(1-\tau_i) = 1/2$.

Let $g(\tau_j) = \prod_{i=2}^{N_s}(1-\tau_i)$. At $\tau_j = 0$, we have $\tau_{k\neq1} = 0$ from (25); then, $g(0) = 1 > 1/2$. Moreover, at $\tau_j = 1$, we have $\tau_{k\neq1} = 1$ from (25); then, $g(1) = 0 < 1/2$. These, together with the fact that $g(\tau_j)$ is a decreasing function of τ_j (due to $\tau_{k\neq1}$ increasing with τ_j from (25)), imply that $f_1(\tau_j) = 0$ has unique solution τ_j^* in (0,1).

Next, to see that the solution to $f_1(\tau_j, c) = f_2(\tau_j)$ is unique, let $f(\tau_j, c) = f_1(\tau_j, c) - f_2(\tau_j)$, which is given by

$$\frac{2}{\prod_{i\in\mathcal{P}\setminus\{1,j\}}(1-\tau_i)}\Big(\frac{1}{cW_j(1-\tau_j)} - \frac{1}{4-W_j\tau_j}\Big) + 1 - \frac{4}{cW_j}$$
$$\equiv g_1(\tau_j)g_2(\tau_j) + 1 - \frac{4}{cW_j}$$

Clearly $g_1(\tau_j)$ is increasing and positive for $\tau_j \in [0, 1)$. Moreover, $g_2(\tau_j)$ is negative since (5) implies the second term is negative, and the hypothesis $W_i > 11$ for all i implies that $cW_j(1-\tau_j) > 4(1-\tau_j) > 4 - W_j\tau_j$. Similarly, g_2 is decreasing because its derivative

$$g_2'(\tau_j) = \frac{1}{cW_j}\frac{1}{(1-\tau_j)^2} - \frac{W_j}{(4-W_j\tau_j)^2}$$
$$< \frac{1}{4(1-\tau_j)^2} - \frac{W_j}{(4-W_j\tau_j)^2} = \frac{(4-W_j)(4-W_j\tau_j^2)}{4(1-\tau_j)^2(4-W_j\tau_j)^2} < 0$$

which uses the fact that $4-W_j\tau_j^2 \geq 4-W_j\tau_j > 0$ by (23) and $4 - W_j < 0$. Thus $f(\tau_j, c)$ is decreasing in τ_j. This implies that the solution to $f_1(\tau_j, c) = f_2(\tau_j)$ is unique.

We will now investigate how this unique solution changes with the action of user 1. When user 1 changes its action, its $CW_{\min}$ (W_1) changes, causing the coefficient c in (20) to change. Let τ_{j1} and τ_{j2} be the solutions to $f(\tau_j, c) = 0$ for $c = c_1$ and $c = c_2 > c_1$, respectively.

It is clear that $f(\tau_j, c)$ is also increasing in c; hence, $f(\tau_{j1}, c_2) > f(\tau_{j1}, c_1) = f(\tau_{j2}, c_2) = 0$. This, together with the fact that $f(\tau_j, c)$ is a decreasing function of τ_j, implies that $\tau_{j1} < \tau_{j2}$. Therefore, when c increases or W_1 increases, τ_j increases and τ_1 decreases.

In particular, c decreases when a_1 changes from B_k to B_{k-1} while a_j remain unchanged; hence, this change decreases τ_j and increases τ_1. ■

D. Theorem 6

Let $\tau_i(a)$, $p_i(a)$ and $W_i(a)$ denote the attempt probability, collision probability and CW_{min} of a player $i \in \mathcal{P}$ under the action profile a. Let j denote any player in $\mathcal{P} \setminus \{1\}$.

Proof: The successful transmission rate per slot of the data user in accordance with each action profiles $a_{(B_1;\cdot)}$ and $a_{(B_{k>1};\cdot)}$, respectively, are given from (3) as follows

$$C_1(a_{(B_{k>1};\cdot)}) = \eta_k\tau_1(a_{(B_{k>1};\cdot)})\big(1-p_1(a_{(B_{k>1};\cdot)})\big)\mathcal{T} \quad (29a)$$
$$C_1(a_{(B_1;\cdot)}) = \tau_1(a_{(B_1;\cdot)})\big(1-p_1(a_{(B_1;\cdot)})\big)\mathcal{T}. \quad (29b)$$

To show $C_1(a_{(B_{k>1};\cdot)}) < C_1(a_{(B_1;\cdot)})$, it's sufficient to show

$$\eta_k\tau_1(a_{(B_{k>1};\cdot)}) > \tau_1(a_{(B_1;\cdot)}) \quad (30a)$$
$$p_1(a_{(B_{k>1};\cdot)}) > p_1(a_{(B_1;\cdot)}). \quad (30b)$$

Those will be proven as follows.

The conditions of this theorem satisfy those of Lemma 7. In the action space $A0$, we partition the cases by the action a_1 of user 1.

Consider $a_1 = B_1$. From (26),

$$\tau_1(a_{(B_1;\cdot)}) = \frac{2}{W_{B_1}}\bigg(2 - \frac{1}{\prod_{i=2}^{N_s}(1-\tau_i(a_{(B_1;\cdot)}))}\bigg). \quad (31a)$$

Otherwise, $a_1 = B_k$. From (26),

$$\tau_1(a_{(B_{k>1};\cdot)}) = \frac{2}{\eta_k W_{B_1}}\bigg(2 - \frac{1}{\prod_{i=2}^{N_s}(1-\tau_i(a_{(B_{k>1};\cdot)}))}\bigg). \quad (31b)$$

When a_1 changes from $B_{k>1}$ to B_1, c in (20) decreases because class $B_{k>1}$ has higher CW_{min} than class B_1. Then, from Lemma 7, for any player $j \neq 1$, we have

$$\tau_j(a_{(B_{k>1};\cdot)}) > \tau_j(a_{(B_1;\cdot)}). \quad (32)$$

From (31) and (32), we obtain (30a) as follows

$$\tau_1(a_{(B_1;\cdot)}) = \eta_k\frac{2}{\eta_k W_{B_1}}\bigg(2 - \frac{1}{\prod_{i=2}^{N_s}(1-\tau_i(a_{(B_1;\cdot)}))}\bigg)$$
$$> \eta_k\frac{2}{\eta_k W_{B_1}}\bigg(2 - \frac{1}{\prod_{i=2}^{N_s}(1-\tau_i(a_{(B_{k>1};\cdot)}))}\bigg)$$
$$= \eta_k\tau_1(a_{(B_{k>1};\cdot)}).$$

Applying (32) to (21) gives (30b). ■

E. Theorem 8

Proof: First, note that $S_1(a_{(B_k;\cdot;B_{k-1};\cdot)}) = S_{s_k}$ and $S_j(a_{(B_k;\cdot;B_{k-1};\cdot)}) = S_{s_{k-1}}$ under the wireless model (2)–(4). Therefore, it is sufficient to show that all ϵ satisfying (7) satisfy $S_{s_k}/S_{s_{k-1}} > 1$, as follows.

Let $\phi(W, p) = W(1-p)/(1-2p)$. With this notation, dividing τ_{s_k} from (2a) by $\tau_{s_{k-1}}$ from (2a), gives

$$\frac{\tau_{s_k}}{\tau_{s_{k-1}}} = \frac{\phi(W_{B_{k-1}}, p_{s_{k-1}}) + 1}{\phi(W_{B_k}, p_{s_k}) + 1}. \quad (33)$$

Moreover, we can apply Lemma 5 since, by hypothesis,

$$W_{B_k} = \frac{\eta_k}{\eta_{k-1}}W_{B_{k-1}} - \epsilon_k > W_{B_{k-1}} > 11$$

Hence $p_{s_k} > p_{s_{k-1}}$ by (2c). Since $W_{B_k} > W_{B_{k-1}}$, this implies $\phi(W_{B_k}, p_{s_k}) > \phi(W_{B_{k-1}}, p_{s_{k-1}})$, whence (33) gives

$$\frac{\tau_{s_k}}{\tau_{s_{k-1}}} > \frac{\phi(W_{B_{k-1}}, p_{s_{k-1}})}{\phi(W_{B_k}, p_{s_k})}. \quad (34)$$

By (3), dividing S_{s_k} from (4) by $S_{s_{k-1}}$ from (4), and then substituting (34) and the definition of ϕ gives

$$\frac{S_{s_k}}{S_{s_{k-1}}} = \frac{\eta_k\tau_{s_k}(1-p_{s_k})}{\eta_{k-1}\tau_{s_{k-1}}(1-p_{s_{k-1}})} > \frac{\eta_k W_{B_{k-1}}(1-2p_{s_k})}{\eta_{k-1}W_{B_k}(1-2p_{s_{k-1}})}. \quad (35)$$

It remains to show that the right hand side exceeds 1.

Dividing $1-\tau_{s_k}$ by $1-\tau_{s_{k-1}}$ with τ_{s_k} and $\tau_{s_{k-1}}$ from (2a) gives

$$\frac{1-\tau_{s_k}}{1-\tau_{s_{k-1}}} = \frac{1 - \dfrac{2}{\phi(W_{B_k}, p_{s_k}) + 1}}{1 - \dfrac{2}{\phi(W_{B_{k-1}}, p_{s_{k-1}}) + 1}}$$
$$< \frac{1 - 2/\phi(W_{B_k}, p_{s_k})}{1 - 2/\phi(W_{B_{k-1}}, p_{s_{k-1}})}$$
$$= \frac{W_{B_k} - 2 - p_{s_k}(W_{B_k} - 4)}{W_{B_{k-1}} - 2 - p_{s_{k-1}}(W_{B_{k-1}} - 4)}\frac{W_{B_{k-1}}}{W_{B_k}}\frac{1-p_{s_{k-1}}}{1-p_{s_k}}. \quad (36)$$

since $\phi(W_{B_k}, p_{s_k}) > \phi(W_{B_{k-1}}, p_{s_{k-1}}) > 1$.

The final factor of (36) cancels with the left hand side by (15), and so the hypothesis $W_{B_k} > 4$ implies

$$1 - 2p_{s_k} >$$
$$1 - 2\frac{W_{B_k} - 2 - \big(W_{B_{k-1}} - 2 - (W_{B_{k-1}} - 4)p_{s_{k-1}}\big)\frac{W_{B_k}}{W_{B_{k-1}}}}{W_{B_k} - 4}$$
$$= \frac{W_{B_k}}{W_{B_{k-1}}}\frac{W_{B_{k-1}} - 4}{W_{B_k} - 4}(1 - 2p_{s_{k-1}}).$$

Substituting this into (35) and using the fact that $1 - 2p_{s_{k-1}} > 0$ we obtain

$$\frac{S_{s_k}}{S_{s_{k-1}}} > \frac{\eta_k W_{B_{k-1}}(1-2p_{s_k})}{\eta_{k-1} W_{B_k}(1-2p_{s_{k-1}})} > \frac{\eta_k}{\eta_{k-1}}\frac{W_{B_{k-1}} - 4}{W_{B_k} - 4}. \quad (37)$$

For $W_{B_k} = \frac{\eta_k}{\eta_{k-1}} W_{B_{k-1}} - \epsilon_k$ with $\epsilon_k \geq 4(\frac{\eta_k}{\eta_{k-1}} - 1)$, the most right hand side of (37) is at least 1, which implies that $S_{s_k} > S_{s_{k-1}}$. ■

F. Lemma 2

Proof: To prove that the attempt probability of a data user reduces when its $CW_{\min}$ increases, we first find the fixed point and then prove its property when $CW_{\min}$ changes.

By hypothesis, we will consider the network with $N_u = 0$ and $N_s = N_{s_k}$. Then, (2) becomes

$$\tau_{s_k} = \frac{2}{W_{B_k}\frac{1-p_{s_k}}{1-2p_{s_k}} + 1} \equiv g_1(p_{s_k}) \quad (38a)$$
$$p_{s_k} = 1 - (1 - \tau_{s_k})^{N_s - 1}. \quad (38b)$$

From (38b),

$$\tau_{s_k} = 1 - (1 - p_{s_k})^{1/(N_s-1)} \equiv g_2(p_{s_k}). \quad (39)$$

The solution of (38) is the solution to $g_1(p_{s_k}) = g_2(p_{s_k})$. Next, we will prove that there exists a solution to $g_1(p_{s_k}) = g_2(p_{s_k})$ and the solution is unique.

First, for finite N_s,

$$g_1(0) = \frac{2}{W_{B_k} + 1} > g_2(0) = 0$$
$$g_1(1/2) = 0 < g_2(1/2) = 1 - (1/2)^{1/(N_s-1)}.$$

This, together with the fact that $g_1(p_{s_k})$ and $g_2(p_{s_k})$ are continuous functions over $[0, \frac{1}{2}]$, implies that there exists solution to $g_1(p_{s_k}) = g_2(p_{s_k})$.

Second, $g_2(p_{s_k})$ is an increasing function of p_{s_k} and $g_1(p_{s_k})$ is a decreasing function of p_{s_k}. Hence, it can be concluded that the solution to $g_1(p_{s_k}) = g_2(p_{s_k})$ is unique.

Next we show how the fixed point changes with $CW_{\min}$. Define $g(p_{s_k}, W_{B_k})$ by

$$g(p_{s_k}, W_{B_k}) = g_1(p_{s_k}) - g_2(p_{s_k}).$$

Let $p_{s_{k1}}$ and $p_{s_{k2}}$ be the solution to $g(p_{s_k}, W_{B_k}) = 0$ at $W_{B_k} = W_{B_{k1}}$ and $W_{B_k} = W_{B_{k2}} > W_{B_{k1}}$, respectively.

It is clear that $g(p_{s_k}, W_{B_k})$ is a decreasing function of W_{B_k}; hence, $g(p_{s_{k2}}, W_{B_{k1}}) > g(p_{s_{k2}}, W_{B_{k2}}) = g(p_{s_{k1}}, W_{B_{k1}}) = 0$. This, together with the fact that $g(p_{s_k}, W_{B_k})$ is a decreasing function of p_{s_k}, implies that $p_{s_{k2}} < p_{s_{k1}}$.

From (38b), $p_{s_{k2}} < p_{s_{k1}}$ implies $\tau_{s_{k2}} < \tau_{s_{k1}}$. This is illustrated in Fig. 8. ■

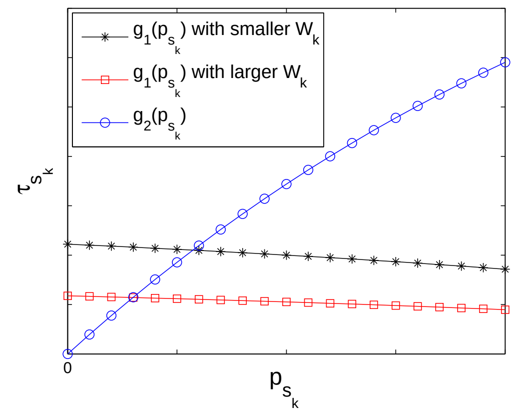


Fig. 8. Graphs of (38a) and (39) at different W_{B_k}.

G. Theorems 1 and 9

Theorems 1 and 9 are immediate corollaries of the following result, with $(M_1, M_2) = (0, 0)$ and $(4, 4)$ respectively.

Lemma 13. *Consider the wireless model (2)–(4) with $N_u = 0$, when all data users use class B_k with $W_{B_k} = \frac{\eta_k}{\eta_{k-1}}(W_{B_{k-1}} - M_1) + M_2$ for constants $M_1 < W_{B_{k-1}}$, $M_2 \geq 0$ and $M_2 \leq M_1$, their throughput per second increases in comparison with using class B_{k-1} .*

Proof: Consider two networks with $N_s > 0$, identical except that one has all data users using class B_{k-1} and the other has all data users using class B_k. Quantities pertaining to the two networks will be designated by subscripts $i \in \{k-1, k\}$.

From (11) and (2c) with users of the same class,

$$\mathbb{E}[Y_i] = \sigma(1-\tau_{s_i})(1-p_{s_i}) + N_s T^s_{s_i}\tau_{s_i}(1-p_{s_i})$$
$$+ \sum_{x\in S} T_x \tau_{s_i}\big((1-\tau_{s_i})^{N_{<x}} - (1-\tau_{s_i})^{N_s-1}\big) \quad (40)$$

where $N_{<x}$ is the number of saturated sources with packets no larger than T_x.

Substituting (40) and (3) into (4) and then dividing numerator and denominator by $\tau_{s_i}(1-p_{s_i})\eta_i$ gives

$$\frac{\mathcal{T}}{S_{s_i}} = \sigma\Big(\frac{1-\tau_{s_i}}{\eta_i\tau_{s_i}}\Big) + T^s_{s_i}\frac{N_s}{\eta_i} + \sum_{x\in S}\frac{T_x}{\eta_i}\Big(\frac{1}{(1-\tau_{s_i})^{N_s-N_{<x}-1}} - 1\Big) \quad (41)$$

To show $S_{s_{k-1}} < S_{s_k}$, it's sufficient to show that the right hand side of (41) is higher for $S_{s_{k-1}}$ than for S_{s_k}. Since $\eta_k > \eta_{k-1}$, it is sufficient that both

$$\frac{1}{(1-\tau_{s_k})^{N_s-N_{<x}-1}} \leq \frac{1}{(1-\tau_{s_{k-1}})^{N_s-N_{<x}-1}}. \quad (42a)$$
$$\sigma\Big(\frac{1-\tau_{s_k}}{\eta_k\tau_{s_k}}\Big) + T^s_{s_k}\frac{N_s}{\eta_k} < \sigma\Big(\frac{1-\tau_{s_{k-1}}}{\eta_{k-1}\tau_{s_{k-1}}}\Big) + T^s_{s_{k-1}}\frac{N_s}{\eta_{k-1}} \quad (42b)$$

1) Proof of (42a): Because the conditions of Lemma 13 satisfy those of Lemma 2, we have

$$\tau_{s_{k-1}} > \tau_{s_k} \qquad p_{s_{k-1}} > p_{s_k}. \quad (43)$$

Since $\tau_{s_{k-1}} > \tau_{s_k}$ by (43), the fact that $(1-\tau_{s_i})^{N_s-N_{<x}-1}$ is non-increasing with the increase of τ_{s_i} establishes (42a).

2) Proof of (42b): Showing (42b) is equivalent to showing the right hand side of (42b) subtracted by the left hand side is greater than 0.

From (12),

$$\Big(\sigma\Big(\frac{1-\tau_{s_{k-1}}}{\eta_{k-1}\tau_{s_{k-1}}}\Big)+T^s_{s_{k-1}}\frac{N_s}{\eta_{k-1}}\Big)-\Big(\sigma\Big(\frac{1-\tau_{s_k}}{\eta_k\tau_{s_k}}\Big)+T^s_{s_k}\frac{N_s}{\eta_k}\Big)$$
$$=\sigma\Big(\frac{1}{\eta_{k-1}\tau_{s_{k-1}}}-\frac{1}{\eta_k\tau_{s_k}}\Big)+\Big(\frac{1}{\eta_{k-1}}-\frac{1}{\eta_k}\Big)(EN_s-\sigma). \tag{44}$$

Since $E>\sigma$ by (12), to show that (44) is greater than 0, it suffices to show

$$\eta_{k-1}\tau_{s_{k-1}}<\eta_k\tau_{s_k} \tag{45}$$

as below.

Multiplying $\tau_{s_{k-1}}$ and τ_{s_k} from (2a) by η_{k-1} and η_k, respectively, gives

$$\frac{2}{\tau_{s_{k-1}}\eta_{k-1}}=\frac{W_{B_{k-1}}}{\eta_{k-1}}\frac{1-p_{s_{k-1}}}{1-2p_{s_{k-1}}}+\frac{1}{\eta_{k-1}} \tag{46}$$

$$\frac{2}{\tau_{s_k}\eta_k}=(\frac{W_{B_{k-1}}-M_1}{\eta_{k-1}}+\frac{M_2}{\eta_k})\frac{1-p_{s_k}}{1-2p_{s_k}}+\frac{1}{\eta_k}. \tag{47}$$

Applying $\eta_k>\eta_{k-1}$ and $M_2\le M_1$ to (46) and (47),

$$\frac{1}{\eta_k}<\frac{1}{\eta_{k-1}}, \tag{48}$$

$$\frac{W_{B_{k-1}}-M_1}{\eta_{k-1}}+\frac{M_2}{\eta_k}<\frac{W_{B_{k-1}}}{\eta_{k-1}}. \tag{49}$$

By (43),

$$\frac{1-p_{s_{k-1}}}{1-2p_{s_{k-1}}}<\frac{1-p_{s_k}}{1-2p_{s_k}}. \tag{50}$$

Substituting those into (46) and (47) implies (45). ∎

H. Lemma 12

Proof: Consider action profiles $a_{(B_k;\cdot;B_{k+i};\cdot)}$ ($k<m$, $i\ge 0$ and $k+i\le m$) and $a_{(B_m;\cdot;B_{k+i};\cdot)}$.

To show (9), we first show that

$$C_j(a_{(B_m;\cdot;B_{k+i};\cdot)})>C_j(a_{(B_k;\cdot;B_{k+i};\cdot)}) \tag{51}$$

as follows.

When a_1 changes from using class B_k to B_m, we have the following from Lemma 7 due to $W^B_k<W^B_m$

$$\tau_j(a_{(B_k;\cdot;B_{k+i};\cdot)})<\tau_j(a_{(B_m;\cdot;B_{k+i};\cdot)}). \tag{52}$$

From (23), p_i is decreasing in τ_i. Then by (52),

$$p_j(a_{(B_k;\cdot;B_{k+i};\cdot)})>p_j(a_{(B_m;\cdot;B_{k+i};\cdot)}). \tag{53}$$

From (3), the successful transmission rates per slot of the data user j under the action profile $a_{(B_h;\cdot;B_{k+i};\cdot)}$ ($h\le m$) is

$$C_j(a_{(B_h;\cdot;B_{k+i};\cdot)})=\eta_{k+i}\tau_j(a_{(B_h;\cdot;B_{k+i};\cdot)}) .(1-p_j(a_{(B_h;\cdot;B_{k+i};\cdot)}))\mathcal{T} \tag{54a}$$

Substituting (52) and (53) into $C_j(a_{(B_k;\cdot;B_{k+i};\cdot)})$ and $C_j(a_{(B_m;\cdot;B_{k+i};\cdot)})$ from (54) gives (51).

Then, applying Theorem 8 in the action profile $a_{(B_k;\cdot;B_{k+i};\cdot)}$ and $a_{(B_m;\cdot;B_{k+i};\cdot)}$ gives

$$C_1(a_{(B_k;\cdot;B_{k+i};\cdot)})\le C_j(a_{(B_k;\cdot;B_{k+i};\cdot)}) \tag{55}$$

$$C_1(a_{(B_m;\cdot;B_{k+i};\cdot)})\ge C_j(a_{(B_m;\cdot;B_{k+i};\cdot)}). \tag{56}$$

From (51), (55), and (56), we have (9). ∎

I. Theorem 11

Proof: Note that the conditions of this theorem satisfy those of Lemma 12.

Consider an action profile with at least one data user using a class other than B_m. Choose the data user using the lowest class among all users under this action profile. Then, according to Lemma 12, that user has incentive to change its action to using class B_m to improve its throughput. Therefore, it can be concluded that no action profile in which at least one data user using lower class than B_m is a Nash equilibrium. ∎

VII. Conclusion

It is important to provide differentiated services, without giving all users the incentive to use a "highest priority" class. This paper has shown through both analysis and simulation that allowing users to adjust $CW_{\min}$ and *TXOP limit* in the same proportion provides service differentiation in WLANs. This scheme improves service for both data and real-time traffic, especially at high load. However, it still provides a slight incentive for data users to use real-time class's parameters. This misalignment of incentives can be removed by increasing $CW_{\min}$ by a slightly smaller factor than the *TXOP limit*. Our incentive adjusted scheme has many advantages over prior proposals: it improves service for both data and real-time traffic and provides the correct incentives for application optimizers, while allowing easy implementation: a single set of 802.11e MAC parameters provides tradeoff between throughput and delay over the range of load studied.

Acknowledgment

This work was supported by Australian Research Council grants DP1095103 and FT0991594.

References

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11^{TM}, 2007.

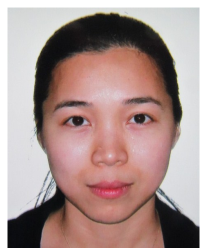
[2] M. Cagalj, S. Ganeriwal, I. Aad, and J.P. Hubaux, "On selfish behavior in CSMA/CA networks," in *Proc. IEEE INFOCOM*, 2005.

[3] L. Galluccio, "A Game-Theoretic Approach to Prioritized Transmission in Wireless CSMA/CA Networks," in IEEE Vehicular Technology Conference, April 2009.

[4] K. Pelechrinis, G. Yan, S. Eidenbenz, S. V. Krishnamurthy, "Detecting Selfish Exploitation of Carrier Sensing in 802.11 Networks," in *Proc. IEEE INFOCOM*, 2009.

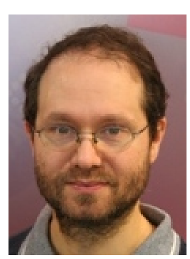
[5] N. Baldo and A. Zanella, "A Game Theoretic evaluation of Rate Adaptation strategies for IEEE 802.11 based Wireless LANs, " in *Proc. VALUETOOLS*, 2009.

[6] P. Chaporkar, A. Proutiere, and B. Radunoviac, “Rate adaptation games in wireless LANs: Nash equilibrium and price of anarchy, ” in *Proc. IEEE INFOCOM*, 2010.
[7] L. Chen and J. Leneutre, “A Game Theoretic Framework of Distributed Power and Rate Control in IEEE 802.11 WLANs, ” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, 2008.
[8] L. Berlemann, G. R. Hiertz, B. H. Walke, and S. Mangold, “Radio Resource Sharing Games: Enabling QoS Support in Unlicensed Bands, ” *IEEE Network*, 2005.
[9] L. Romdhani, N. Qiang, and T. Turletti, “Adaptive EDCF: enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks, ” in *Proc. IEEE WCNC*, 2003.
[10] J. Lv, X. Zhang, and X. Han, “A Novel Dynamic Tuning of the Contention Window (CW) for IEEE 802.11e Enhanced Distributed Control Function, ” in *Proc. Networked Computing and Advanced Information Management*, 2008.
[11] T. Nilsson and J. Farooq, “A Novel MAC Scheme for Solving the QoS Parameter Adjustment Problem in IEEE 802.11e EDCA, ” in *Proc. WoWMoM*, 2008.
[12] T. Jun, X. Xing, C. Zhi-lan, Y. Zhi-wei, and Z. Zhi, “Improving Throughput for Heterogeneous Traffic in IEEE802.11e EDCA, ” in *Proc. IEEE Workshop on IDAACS*, 2009.
[13] N. H. Vaidya, A. Dugar, S. Gupta and P. Bahl, “Distributed fair scheduling in a wireless LAN,” *IEEE Trans. Mobile Computing*, vol. 4, 2005.
[14] P. Nuggehalli, M. Sarkar, K. Kulkarni, and R. R. Rao, “Game-theoretic Analysis of QoS in Wireless MAC,” in *Proc. INFOCOM* 2008.
[15] J. Price, P. Nuggehalli, and T. Javidi, “Incentive Compatible MAC-Layer QoS Design,” In *Proc. CCNC* 2008.
[16] C.-T. Chou, S. N. Shankar and K. G. Shin, “Achieving per-stream QoS with distributed airtime allocation and admission control in IEEE 802.11e wireless LANs” in *Proc. INFOCOM* pp. 1584–1595, 2005.
[17] M. H. Cheung, A. H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, “Random Access Protocols for WLANs based on mechanism design,” in *Proc. IEEE Int. Conf. Commun.*, 2009.
[18] J. Price, P. Nuggehalli, and T. Javidi, “Pricing and QoS in Wireless Random Access Networks,” in *Proc. GLOBECOM* 2008.
[19] P. Hurley and J.Y. Le Boudec, “ABE: Providing a low-delay service within best effort,” *IEEE Network*, vol. 15, no. 3, pp. 60–69, 2001.
[20] M. Karsten, Y. Lin, and K. Larson, “Incentive-Compatible Differentiated Scheduling,” in *Proc. HotNets IV*, 2005.
[21] B. Gaidioz and P. Primet, “EDS: A new scalable service differentiation architecture for internet,” In *Proc. IEEE ISCC*, pp. 777–782, July 2002.
[22] V. Firoiu, X. Zhang, and Y. Guo, “Best Effort Differentiated Services: Trade-off Service Differentiation for Elastic Applications,” In *Proc. IEEE ICT*, 2001.
[23] G. Tan and J. Guttag, “The 802.11 MAC protocol leads to inefficient equilibria,” In *Proc. IEEE INFOCOM*, 2005.
[24] S. H. Nguyen, H. L. Vu and L. L. H. Andrew, “Performance Analysis of IEEE 802.11 WLANs With Saturated and Unsaturated Sources, ” *IEEE Trans. Vehicular Technology*, vol. 61, no. 1, pp. 333–345, 2012.
[25] V. Ramaiyan, A. Kumar, and E. Altman, “Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability, ” *IEEE/ACM Trans. Networking*, vol. 16, no. 5, pp. 1080 – 1093 , 2008.
[26] P. K. Dutta, *Strategies and games: theory and practice*. The MIT Press, 1999.
[27] D. Fudenberg and J. Tirole, *Game Theory*. The MIT Press, 1991.
[28] “The network simulator ns-2,” Available at ⟨http://www.isi.edu/nsnam/ns/⟩.
[29] S. Wietholter and C. Hoene, “An IEEE 802.11e EDCF and CFB simulation model for ns-2,” Available at ⟨http://www.tkn.tu-berlin.de/research/802.11e_ns2/⟩.



Suong H. Nguyen received B.Sc. degree and M.Sc. from the Post and Telecommunications Institute of Technology, Vietnam in 2000 and La Trobe University, Australia in 2007, respectively. She is currently Ph.D. student in Swinburne University of Technology, Australia. Her research interest includes wireless communication and optical transmission system.

Hai L. Vu (S'97M'98-SM'06) received the B.Sc./M.Sc. and Ph.D. degrees in electrical engineering from the Technical University of Budapest, Budapest, Hungary, in 1994 and 1999, respectively. From 1994 to 2000, he was a Research Engineer with Siemens AG, Hungary. During 2000-2005, he was with the Department of Electrical and Electronic Engineering, University of Melbourne, Australia. In 2005, he joined Swinburne University of Technology where he is currently an Associate Professor. Dr. Vu has authored or coauthored over 100 scientific journals and conference papers. His research interests include performance analysis and design of wireless data networks, and stochastic optimization with applications to Intelligent Transport Systems (ITS) and SmartGrid.



Lachlan Andrew (M'97-SM'05) received the B.Sc., B.E. and Ph.D. degrees in 1992, 1993, and 1997, from the University of Melbourne, Australia. Since 2008, he has been an associate professor at Swinburne University of Technology, Australia, and since 2010 he has been an ARC Future Fellow. From 2005 to 2008, he was a senior research engineer in the Department of Computer Science at Caltech. Prior to that, he was a senior research fellow at the University of Melbourne and a lecturer at RMIT, Australia. His research interests include energy-efficient networking and performance analysis of resource allocation algorithms. He was co-recipient of the best paper award at IEEE INFOCOM 2011 and IEEE MASS 2007. He is a member of the ACM.

APPENDIX A
PROOF OF THEOREM 3

Proof: We first prove Claim (T3-1) that under the proportional scheme (1), the throughput per slot of a saturated source using class B_k increases when η_k increases. Then, we prove Claim (T3-2) that the collision probability of unsaturated sources decreases with the increase of $\eta_k \geq 1$.

Although the scenario is simple, the proof is complicated by the mutual independence between attempt probabilities and collision probabilities in the fixed point, their dependence on the scale η_k, and the dependence of the slot time $\mathbb{E}[Y]$ on η_k.

1) Proof of Claim (T3-1): Substituting (1), (5) into (3) gives

$$C_{s_k} = \frac{2}{W_{B_1}}(1-2p_{s_k})\mathcal{T}$$

which is decreasing in p_{s_k}. Thus, to prove Claim (T3-1), it is sufficient to show that $dp_{s_k}/d\eta < 0$. To prove $dp_{s_k}/d\eta < 0$, we first find a closed-form expression of $\frac{dp_{s_k}}{d\eta_k}$ from (58). This can be achieved by solving two linear equations of $dp_{s_k}/d\eta$ and $d\tau_{s_k}/d\eta$. The first equation is $dp_{s_k}/d\eta$ as a linear function of $\frac{d\tau_{s_k}}{d\eta_k}$ and the second one is $\frac{d\tau_{s_k}}{d\eta_k}$ as a linear function of $dp_{s_k}/d\eta$, which are determined as follows.

Firstly, we find $dp_{s_k}/d\eta$ as a linear function of $\frac{d\tau_{s_k}}{d\eta_k}$. Recall that $N_u = N_{s_k} = N_s = 1$, whence by (2c), $p_{s_k} = \tau_u$ and $p_u = \tau_{s_k}$. Hence by (2b),

$$p_{s_k} = \lambda_u \mathbb{E}[Y]\frac{1}{1-\tau_{s_k}} \tag{57}$$

with $\mathbb{E}[Y]$ given by (11).

Taking the derivative of (57) with respect to η_k gives

$$\frac{dp_{s_k}}{d\eta_k} = \frac{\lambda_u}{(1-\tau_{s_k})^2}\Big(\frac{d\mathbb{E}[Y]}{d\eta_k}(1-\tau_{s_k}) + \mathbb{E}[Y]\frac{d\tau_{s_k}}{d\eta_k}\Big). \tag{58}$$

To have $dp_{s_k}/d\eta$ as a linear function of $\frac{d\tau_{s_k}}{d\eta_k}$ from (58), we now express $\frac{d\mathbb{E}[Y]}{d\eta_k}$ in terms of $d\tau_{s_k}/d\eta_k$. From (5), $W_{s_k} = \eta_k W_{B_1}$ implies

$$p_{s_k} = 1 - \frac{2/W_{B_1}}{4/W_{B_1} - \eta_k\tau_{s_k}}. \tag{59}$$

Substituting $N_u = N_s = N_{s_k} = 1$ by hypothesis, (2c), $p_{s_k} = \tau_u$, $p_u = \tau_{s_k}$, and (59) into (11) gives

$$\begin{aligned}\mathbb{E}[Y] &= (\sigma - T_u)(1-\tau_{s_k})\frac{2/W_{B_1}}{4/W_{B_1}-\eta_k\tau_{s_k}} + T_c \\ &+ (T_u - T_c)(1-\tau_{s_k}) + (T^s_{s_k} - T_c)\tau_{s_k}\frac{2/W_{B_1}}{4/W_{B_1}-\eta_k\tau_{s_k}}.\end{aligned} \tag{60}$$

where $T_c = \max(T_u, T_s)$. This shows that $\mathbb{E}[Y]$ is a function of τ_{s_k} and $T^s_{s_k}$, both of which depend on η_k.

By (13), T_u and T_c are independent of η_k and $dT^s_s/d\eta_k = \mathcal{T}$ from (12). Besides, τ_{s_k} is also a function of η_k. Then, taking derivative of (60) and then substituting (12) into the obtained equation gives

TABLE V
MATH EXPRESSION OF SYMBOLS IN THEOREM 3.

Symbol	Expression
K_1	$(\sigma - T_u)(2/W_{B_1})(\eta_k - 4/W_{B_1})$ $-(T_u - T_c)(4/W_{B_1} - \eta_k\tau_{s_k})^2 + (T^s_{s_k} - T_c)(8/W^2_{B_1})$
K_2	$(2/W_{B_1})\tau_{s_k}\Big((\sigma - T_u)(1-\tau_{s_k})$ $+\mathcal{T}(4/W_{B_1}) + (E - T_c)\tau_{s_k}\Big)$
L_1	$1 + \frac{\lambda_u}{(1-\tau_{s_k})^2}\Big(\frac{K_1(1-\tau_{s_k})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} + \mathbb{E}[Y]\Big)$ $\cdot\frac{(2/W_{B_1})}{\eta_k}\frac{1}{(1-p_{s_k})^2}$
L_2	$\frac{K_2(1-\tau_{s_k})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} - \Big(\frac{K_1(1-\tau_{s_k})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} + \mathbb{E}[Y]\Big)\frac{\tau_{s_k}}{\eta_k}$
H_1	$\frac{1}{(4/W_{B_1}-\eta_k\tau_{s_k})^2}\Big(-(2/W_{B_1})\tau_{s_k} - \frac{\lambda_u K_2}{1-\tau_{s_k}}\Big)$
H_2	$\frac{\eta_k(2/W_{B_1})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2}$ $+\frac{\lambda_u}{1-\tau_{s_k}}\Big(\frac{\mathbb{E}[Y]}{1-\tau_{s_k}} + \frac{K_1}{(4/W_{B_1}-\eta_k\tau_{s_k})^2}\Big)$

$$\begin{aligned}\frac{d\mathbb{E}[Y]}{d\eta_k} &= \frac{(\sigma - T_u)(2/W_{B_1})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2}\Big(\big(\eta_k - \frac{4}{W_{B_1}}\big)\frac{d\tau_{s_k}}{d\eta_k} + \tau_{s_k} - \tau^2_{s_k}\Big) \\ &- (T_u - T_c)\frac{d\tau_{s_k}}{d\eta_k} + \frac{(2/W_{B_1})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} \\ &\cdot\Big((T^s_{s_k} - T_c)(4/W_{B_1})\frac{d\tau_{s_k}}{d\eta_k} + \mathcal{T}(4/W_{B_1})\tau_{s_k} + (E - T_c)\tau^2_{s_k}\Big) \\ &= \frac{1}{(4/W_{B_1}-\eta_k\tau_{s_k})^2}\Big(K_1\frac{d\tau_{s_k}}{d\eta_k} + K_2\Big)\end{aligned} \tag{61}$$

where K_1 and K_2 are given in Table V.

Then, substituting (61) into (58) gives

$$\begin{aligned}\frac{dp_{s_k}}{d\eta_k} =& \frac{\lambda_u}{(1-\tau_{s_k})^2}\Big(\frac{K_2(1-\tau_{s_k})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} \\ &+ \Big(\frac{K_1(1-\tau_{s_k})}{(4/W_{B_1}-\eta_k\tau_{s_k})^2} + \mathbb{E}[Y]\Big)\frac{d\tau_{s_k}}{d\eta_k}\Big).\end{aligned} \tag{62}$$

Secondly, we find $\frac{d\tau_{s_k}}{d\eta_k}$ as a linear function of $dp_{s_k}/d\eta$. Substituting $W_{s_k} = \eta_k W_{B_1}$ into (5) gives

$$\tau_{s_k} = \frac{(2/W_{B_1})}{\eta_k}\Big(2 - \frac{1}{1-p_{s_k}}\Big) \tag{63}$$

Then, differentiating (63) with respect to η_k gives

$$\begin{aligned}\frac{d\tau_{s_k}}{d\eta_k} &= -\frac{1}{\eta_k}\frac{(2/W_{B_1})}{\eta_k}\Big(2 - \frac{1}{1-p_{s_k}}\Big) - \frac{(2/W_{B_1})}{\eta_k(1-p_{s_k})^2}\frac{dp_{s_k}}{d\eta_k} \\ &= -\frac{1}{\eta_k}\tau_{s_k} - \frac{2/W_{B_1}}{\eta_k}\frac{1}{(1-p_{s_k})^2}\frac{dp_{s_k}}{d\eta_k}.\end{aligned} \tag{64}$$

Note that the last expression uses (63).

Solving two linear equations (62) and (64) gives

$$\frac{dp_{s_k}}{d\eta_k}L_1 = \frac{\lambda_u}{(1-\tau_{s_k})^2}L_2 \tag{65}$$

where L_1 and L_2 are given in Table V. From (65), to show $dp_{s_k}/d\eta_k < 0$, it is sufficient to show $L_1 > 0$ and $L_2 < 0$ as follows.

First, we show $L_1 > 0$. From (59),

$$\frac{1}{(1-p_{s_k})^2} = \frac{(4/W_{B_1} - \eta_k \tau_{s_k})^2}{(2/W_{B_1})^2} = \frac{J}{(2/W_{B_1})^2}. \quad (66)$$

Substituting (60), K_1 from Table V, and (66) into L_1 from Table V gives

$$L_1 = 1 + \lambda_u \Big((\sigma - T_u) + T_c \frac{(4/W_{B_1} - \eta_k \tau_{s_k})^2}{(1-\tau_{s_k})^2} \frac{1}{(2/W_{B_1})\eta_k} + (T^s_{s_k} - T_c) \frac{(4/W_{B_1} - \eta_k \tau_{s_k}^2)}{(1-\tau_{s_k})^2 \eta_k} \Big) \quad (67)$$

Since $\lambda_u T_u \leq 1$ by hypothesis, $T^s_{s_k} \geq T_c$ by (12) and (13), and

$$4/W_{B_1} - \eta_k \tau_{s_k}^2 > 4/W_{B_1} - \eta_k \tau_{s_k} > 0 \quad (68)$$

from (59), we have $L_1 > 0$.

Next, we show $L_2 < 0$. Substituting (12), (60), K_1 and K_2 from Table V into L_2 from Table V gives

$$L_2 = -\frac{(2/W_{B_1})\tau_{s_k}}{4/W_{B_1} - \eta_k \tau_{s_k}} \Big(\tau_{s_k} \mathcal{T} + \frac{E}{\eta_k} \Big) + \Big(\frac{\tau_{s_k}}{\eta_k} \Big) \Big(\frac{-2/W_{B_1} + \eta_k \tau_{s_k}}{4/W_{B_1} - \eta_k \tau_{s_k}} \Big) T_c \quad (69)$$

From (5), we have $\tau_{s_k} < \frac{2}{\eta_k W_{B_1}}$ due to $p_{s_k} \in (0,1)$. Then,

$$-2/W_{B_1} + \eta_k \tau_{s_k} < -2/W_{B_1} + \eta_k \frac{2}{\eta_k W_{B_1}} = 0 \quad (70)$$

This, together with $\mathcal{T} > 0$, $E > 0$ and (68), implies that $L_2 < 0$.

2) Proof of Claim (T3-2): By (2c), $p_u = \tau_{s_k}$, it is sufficient to show that τ_{s_k} decreases when η_k increases. We first find a closed form expression of $d\tau_{s_k}/d\eta_k$ and then prove it to be less than 0. Recall from Section A-1 that the closed form of $d\tau_{s_k}/d\eta_k$ can be found by solving two linear equations (62) and (64), which gives

$$H_1 = \frac{d\tau_{s_k}}{d\eta_k} H_2 \quad (71)$$

where H_1 and H_2 are given in Table V. From (71), to show $d\tau_{s_k}/d\eta_k < 0$, it is sufficient to prove that $H_1 < 0$ and $H_2 > 0$ as follows.

First, we show $H_1 < 0$. Substituting K_2 from Table V into H_1 from Table V gives

$$H_1 = -\frac{(2/W_{B_1})\tau_{s_k}}{J} \Bigg(1 + \lambda_u(\sigma - T_u) + \frac{\lambda_u}{1-\tau_{s_k}} \Big((4/W_{B_1})\mathcal{T} + (E - T_c)\tau_{s_k} \Big) \Bigg) \quad (72)$$

Since $\lambda_u T_u < 1$ and $\tau_{s_k} \in (0,1)$ by hypothesis, to show $H_1 < 0$ it is sufficient to show that $(4/W_{B_1})\mathcal{T} + (E - T_c)\tau_{s_k} > 0$ as follows. From (5), we have $\tau_{s_k} < 2/(\eta_k W_{B_1})$ due to $p_{s_k} \in (0,1)$ and $T_c - E > 0$ by (13). Then,

$$(4/W_{B_1})\mathcal{T} + (E - T_c)\tau_{s_k} > (4/W_{B_1})\mathcal{T} - (T_c - E)\frac{(2/W_{B_1})}{\eta_k} = (2/W_{B_1})\Big(2\mathcal{T} - \frac{T_c}{\eta_k} + \frac{E}{\eta_k} \Big) \quad (73)$$

Since $T_c = \max(T_u, T_s) < 2\mathcal{T}$ by hypothesis, the left hand side of the inequality (73) is greater than 0, which proves that $H_1 < 0$.

Second, we show $H_2 > 0$. Substituting (60) and K_1 from Table V into H_2 from Table V gives

$$H_2 = \frac{(2/W_{B_1})\eta_k}{(4/W_{B_1} - \eta_k \tau_{s_k})^2} \big(1 + \lambda_u(\sigma - T_u) \big) + \frac{\lambda_u}{(1-\tau_{s_k})^2} \Big(T_c + (T^s_{s_k} - T_c) \frac{(2/W_{B_1})(4/W_{B_1} - \eta_k \tau_{s_k}^2)}{(4/W_{B_1} - \eta_k \tau_{s_k})^2} \Big) \quad (74)$$

Moreover, since $\lambda_u T_u \leq 1$ by hypothesis, $T^s_{s_k} \geq T_c$ by (12) and (13), and (68), it follows from (74) that $H_2 > 0$. ■

APPENDIX B
ADDITIONAL SIMULATION RESULTS

A. Wider range of loads

This section extends Section V-B2, by continuing the numerical study of PIA using $\eta = 2$ and $\epsilon = \epsilon^0 = 4$ under 802.11g for a wider range of traffic loads. Except as noted below, the setting is the same as Section V-B2.

1) Scenario 1: The following simulation results show that when $\eta = 2$, PIA provides better service for both traffic types than is the case when there is no service differentiation ($\eta = 1$).

The ratio of the throughput of a data user under the PIA scheme at $\eta = 2$ to that of the case of no service differentiation ($\eta = 1$) is shown in Fig. 9(a) and the corresponding ratio of the mean delay of an unsaturated user is shown in Fig. 9(b), as functions of the number of unsaturated users N_u for different N_s.

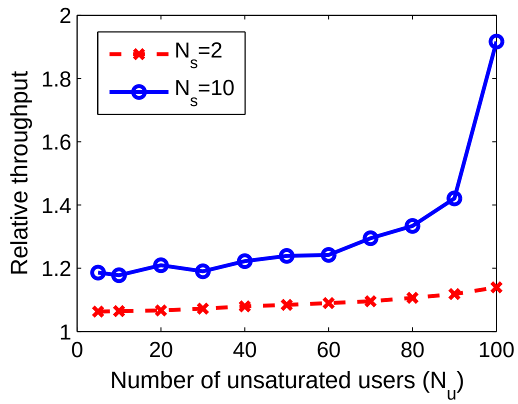
Fig. 9(a) shows that when traffic load increases (N_s and/or N_u increases), the improvement in throughput under PIA in comparison with the case of no service differentiation increases. A similar trend is visible in the delay performance of an unsaturated user shown in Fig. 9(b). When traffic load increases, the delay performance under PIA becomes better than that under no service differentiation.

Note that under light load, the original 802.11 without service differentiation is good enough for both data users and realtime users; although PIA does not help much, help is not necessary, and PIA does not hurt. However, under heavy load where the original 802.11 needs help, our proposed scheme provide significant improvement for both types of traffic.

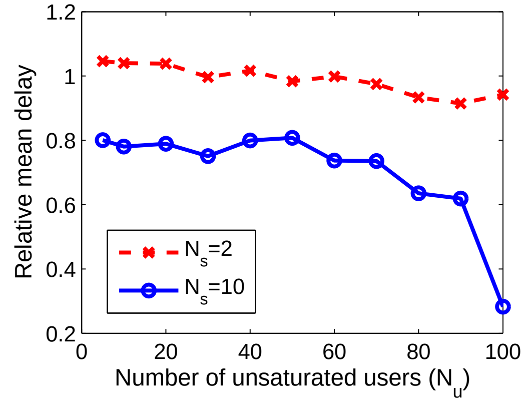
2) Scenario 2: So far, we have only considered the incentives facing saturated data users. Now we consider the incentive of realtime users in choosing a class.

We consider a network with 5 saturated users and 10 unsaturated users. Among the 10 unsaturated users, we tag a particular user and change its arrival rate in a wide range from 10 packets/s to 210 packets/s, while keeping the arrival rate of the other 9 unsaturated users at 40 packets/s. Then, we investigate how the throughput of the tagged user changes with its arrival rate when it uses classes B_1 and B_2, respectively.

Fig. 10 shows that when traffic load (e.g., the arrival rate of the tagged user increases) is not very high, the mean delay of the tagged "unsaturated" user is lower if it uses realtime



(a)



(b)

Fig. 9. Ratio of throughput of a data user under PIA at $\eta = 2$ to that under no service differentiation ($\eta = 1$) and the corresponding ratio of mean delay of a real-time user as a function of the number of realtime users. ($\lambda = 9$ packets/s, $l_{sat} = 1200$ bytes, $l_{nonsat} = 100$ bytes, $\eta = 2$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1}$.)

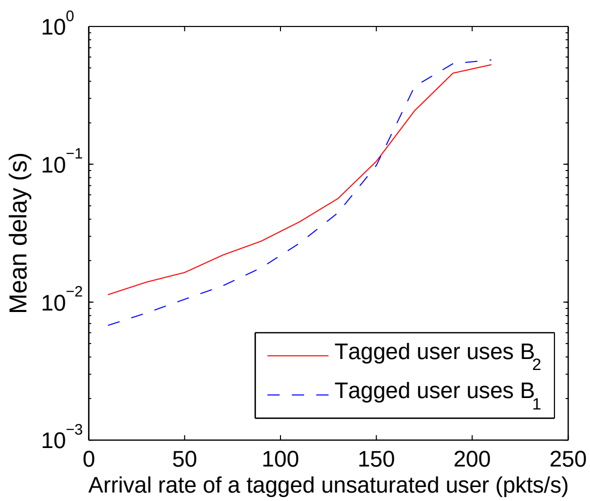


Fig. 10. Mean delay of of the tagged "unsaturated user" as a function of its arrival rate. ($\lambda_{u \neq 1} = 40$ pkts/s, $l_{sat} = 1200$B, $l_{nonsat} = 100$B, $N_u = 10$, $N_s = 5$, $W_{B_1} = 32$, $W_{B_2} = \eta W_{B_1} - \epsilon^0$, $\eta = 2$.)

class B_1, as expected. This is because the tagged user has negligible queueing at that traffic load. However, when traffic load reaches a certain threshold (e.g., the arrival rate of the tagged user is around 150 pkts/s), using the "bulk data" class B_2 gives the tagged "unsaturated" user lower delay. This is because the station incurs non-negligible queuing, and is no longer always unsaturated. This means that our previous assessments of the throughput of class $B_{k>1}$ assuming that realtime users use the realtime class B_1 may not reflect the realtime users' actual choice. However, it is conservative because class B_1 causes the most collisions; if a realtime users change to using the "bulk data" class at high traffic load, this not only improves its delay but also helps data users due to the reduction in collisions.

Appendix C
Implication of Multiple Sources per Station

The MAC model can be used to described many situations in which a single station has multiple sources of data. For example, if the station has two saturated sources using the same access class, then the aggregate throughput is the same as if it had a single saturated source; the fraction of capacity going to each source is beyond the control of the MAC. If the station has two sources using different ACs, then it can be modelled as two separate stations, since each AC operates independently, with its own backoff counters.

Similarly, the game model remains appropriate in the (typical) case that each source does not base its choice on the class chosen by other sources on the same station. That occurs, for example, when an application has the choice of class hard-coded based on performance measurements made by the application writer. Since the fraction of time that typical wireless node is actively transmitting is small, the application should be designed on the assumption that it is the only active source.

If applications dynamically choose the class based on the choices of other sources on the same station, the situation becomes more complicated. Consider a station in which one saturated source has chosen to use the class m with the largest TXOP, giving throughput S. If another saturated source also chooses class m then both will get throughput $S/2$. However, if it chooses class $m-1$, which uses a different AC, then the independence between ACs means that it will achieve the same throughput as if it was the only source at that station, which is only slightly below S. That means that it is no longer a Nash equilibrium for all saturated sources to choose class m. Studying this situation is an interesting direction for future work, although its practical relevance is reduced by the fact that most applications will choose a suitable class at design time rather than run time.